GENETICS, GENOMICS AND BREEDING OF VEGETABLE BRASSICAS
GENETICS, GENOMICS
AND BREEDING OF
VEGETABLE BRASSICAS

Editors
Jan Sadowski
Department of Biotechnology
Institute of Molecular Biology and Biotechnology
Faculty of Biology
Adam Mickiewicz University
Poznań, Poland
and
Laboratory of Functional Genomics
Institute of Plant Genetics
Polish Academy of Sciences
Poznań, Poland

Chittaranjan Kole
Department of Genetics and Biochemistry
Clemson University
Clemson, SC
USA
Dedication

The authors dedicate this monography to Professor César Gómez-Campo, who recently passed away and whose knowledge and remarkable work have been a great inspiration for the Brassica community.
Preface to the Series

Genetics, genomics and breeding has emerged as three overlapping and complimentary disciplines for comprehensive and fine-scale analysis of plant genomes and their precise and rapid improvement. While genetics and plant breeding have contributed enormously towards several new concepts and strategies for elucidation of plant genes and genomes as well as development of a huge number of crop varieties with desirable traits, genomics has depicted the chemical nature of genes, gene products and genomes and also provided additional resources for crop improvement.

In today’s world, teaching, research, funding, regulation and utilization of plant genetics, genomics and breeding essentially require thorough understanding of their components including classical, biochemical, cytological and molecular genetics; and traditional, molecular, transgenic and genomics-assisted breeding. There are several book volumes and reviews available that cover individually or in combination of a few of these components for the major plants or plant groups; and also on the concepts and strategies for these individual components with examples drawn mainly from the major plants. Therefore, we planned to fill an existing gap with individual book volumes dedicated to the leading crop and model plants with comprehensive deliberations on all the classical, advanced and modern concepts of depiction and improvement of genomes. The success stories and limitations in the different plant species, crop or model, must vary; however, we have tried to include a more or less general outline of the contents of the chapters of the volumes to maintain uniformity as far as possible.

Often genetics, genomics and plant breeding and particularly their complimentary and supplementary disciplines are studied and practiced by people who do not have, and reasonably so, the basic understanding of biology of the plants for which they are contributing. A general description of the plants and their botany would surely instill more interest among them on the plant species they are working for and therefore we presented lucid details on the economic and/or academic importance of the plant(s); historical information on geographical origin and distribution; botanical origin and evolution; available germplasms and gene pools, and genetic and cytogenetic stocks as genetic, genomic and breeding resources; and
basic information on taxonomy, habit, habitat, morphology, karyotype, ploidy level and genome size, etc.

Classical genetics and traditional breeding have contributed enormously even by employing the phenotype-to-genotype approach. We included detailed descriptions on these classical efforts such as genetic mapping using morphological, cytological and isozyme markers; and achievements of conventional breeding for desirable and against undesirable traits. Employment of the in vitro culture techniques such as micro- and megaspore culture, and somatic mutation and hybridization, has also been enumerated. In addition, an assessment of the achievements and limitations of the basic genetics and conventional breeding efforts has been presented.

It is a hard truth that in many instances we depend too much on a few advanced technologies, we are trained in, for creating and using novel or alien genes but forget the infinite wealth of desirable genes in the indigenous cultivars and wild allied species besides the available germplasms in national and international institutes or centers. Exploring as broad as possible natural genetic diversity not only provides information on availability of target donor genes but also on genetically divergent genotypes, botanical varieties, subspecies, species and even genera to be used as potential parents in crosses to realize optimum genetic polymorphism required for mapping and breeding. Genetic divergence has been evaluated using the available tools at a particular point of time. We included discussions on phenotype-based strategies employing morphological markers, genotype-based strategies employing molecular markers; the statistical procedures utilized; their utilities for evaluation of genetic divergence among genotypes, local landraces, species and genera, and also on the effects of breeding pedigrees and geographical locations on the degree of genetic diversity.

Association mapping using molecular markers is a recent strategy to utilize the natural genetic variability to detect marker-trait association and to validate the genomic locations of genes, particularly those controlling the quantitative traits. Association mapping has been employed effectively in genetic studies in human and other animal models and those have inspired the plant scientists to take advantage of this tool. We included examples of its use and implication in some of the volumes that devote to the plants for which this technique has been successfully employed for assessment of the degree of linkage disequilibrium related to a particular gene or genome, and for germplasm enhancement.

Genetic linkage mapping using molecular markers have been discussed in many books, reviews and book series. However, in this series, genetic mapping has been discussed at length with more elaborations and examples on diverse markers including the anonymous type 2 markers such as RFLPs, RAPDs, AFLPs, etc. and the gene-specific type 1 markers such as EST-SSRs, SNPs, etc.; various mapping populations including $F_2$, backcross,
recombinant inbred, doubled haploid, near-isogenic and pseudotestcross; computer software including MapMaker, JoinMap, etc. used; and different types of genetic maps including preliminary, high-resolution, high-density, saturated, reference, consensus and integrated developed so far.

Mapping of simply inherited traits and quantitative traits controlled by oligogenes and polygenes, respectively has been deliberated in the earlier literature crop-wise or crop group-wise. However, more detailed information on mapping or tagging oligogenes by linkage mapping or bulked segregant analysis, mapping polygenes by QTL analysis, and different computer software employed such as MapMaker, JoinMap, QTL Cartographer, Map Manager, etc. for these purposes have been discussed at more depth in the present volumes.

The strategies and achievements of marker-assisted or molecular breeding have been discussed in a few books and reviews earlier. However, those mostly deliberated on the general aspects with examples drawn mainly from major plants. In this series, we included comprehensive descriptions on the use of molecular markers for germplasm characterization, detection and maintenance of distinctiveness, uniformity and stability of genotypes, introgression and pyramiding of genes. We have also included elucidations on the strategies and achievements of transgenic breeding for developing genotypes particularly with resistance to herbicide, biotic and abiotic stresses; for biofuel production, biopharming, phytoremediation; and also for producing resources for functional genomics.

A number of desirable genes and QTLs have been cloned in plants since 1992 and 2000, respectively using different strategies, mainly positional cloning and transposon tagging. We included enumeration of these and other strategies for isolation of genes and QTLs, testing of their expression and their effective utilization in the relevant volumes.

Physical maps and integrated physical-genetic maps are now available in most of the leading crop and model plants owing mainly to the BAC, YAC, EST and cDNA libraries. Similar libraries and other required genomic resources have also been developed for the remaining crops. We have devoted a section on the library development and sequencing of these resources; detection, validation and utilization of gene-based molecular markers; and impact of new generation sequencing technologies on structural genomics.

As mentioned earlier, whole genome sequencing has been completed in one model plant (Arabidopsis) and seven economic plants (rice, poplar, peach, papaya, grapes, soybean and sorghum) and is progressing in an array of model and economic plants. Advent of massively parallel DNA sequencing using 454-pyrosequencing, Solexa Genome Analyzer, SOLiD system, Heliscope and SMRT have facilitated whole genome sequencing in many other plants more rapidly, cheaply and precisely. We have included
extensive coverage on the level (national or international) of collaboration and the strategies and status of whole genome sequencing in plants for which sequencing efforts have been completed or are progressing currently. We have also included critical assessment of the impact of these genome initiatives in the respective volumes.

Comparative genome mapping based on molecular markers and map positions of genes and QTLs practiced during the last two decades of the last century provided answers to many basic questions related to evolution, origin and phylogenetic relationship of close plant taxa. Enrichment of genomic resources has reinforced the study of genome homology and synteny of genes among plants not only in the same family but also of taxonomically distant families. Comparative genomics is not only delivering answers to the questions of academic interest but also providing many candidate genes for plant genetic improvement.

The ‘central dogma’ enunciated in 1958 provided a simple picture of gene function—gene to mRNA to transcripts to proteins (enzymes) to metabolites. The enormous amount of information generated on characterization of transcripts, proteins and metabolites now have led to the emergence of individual disciplines including functional genomics, transcriptomics, proteomics and metabolomics. Although all of them ultimately strengthen the analysis and improvement of a genome, they deserve individual deliberations for each plant species. For example, microarrays, SAGE, MPSS for transcriptome analysis; and 2D gel electrophoresis, MALDI, NMR, MS for proteomics and metabolomics studies require elaboration. Besides transcriptome, proteome or metabolome QTL mapping and application of transcriptomics, proteomics and metabolomics in genomics-assisted breeding are frontier fields now. We included discussions on them in the relevant volumes.

The databases for storage, search and utilization on the genomes, genes, gene products and their sequences are growing enormously in each second and they require robust bioinformatics tools plant-wise and purpose-wise. We included a section on databases on the gene and genomes, gene expression, comparative genomes, molecular marker and genetic maps, protein and metabolomes, and their integration.

Notwithstanding the progress made so far, each crop or model plant species requires more pragmatic retrospect. For the model plants we need to answer how much they have been utilized to answer the basic questions of genetics and genomics as compared to other wild and domesticated species. For the economic plants we need to answer as to whether they have been genetically tailored perfectly for expanded geographical regions and current requirements for green fuel, plant-based bioproducts and for improvements of ecology and environment. These futuristic explanations have been addressed finally in the volumes.
We are aware of exclusions of some plants for which we have comprehensive compilations on genetics, genomics and breeding in hard copy or digital format and also some other plants which will have enough achievements to claim for individual book volume only in distant future. However, we feel satisfied that we could present comprehensive deliberations on genetics, genomics and breeding of 30 model and economic plants, and their groups in a few cases, in this series. I personally feel also happy that I could work with many internationally celebrated scientists who edited the book volumes on the leading plants and plant groups and included chapters authored by many scientists reputed globally for their contributions on the concerned plant or plant group.

We paid serious attention to reviewing, revising and updating of the manuscripts of all the chapters of this book series, but some technical and formatting mistakes will remain for sure. As the series editor, I take complete responsibility for all these mistakes and will look forward to the readers for corrections of these mistakes and also for their suggestions for further improvement of the volumes and the series so that future editions can serve better the purposes of the students, scientists, industries, and the society of this and future generations.

Science publishers, Inc. has been serving the requirements of science and society for a long time with publications of books devoted to advanced concepts, strategies, tools, methodologies and achievements of various science disciplines. Myself as the editor and also on behalf of the volume editors, chapter authors and the ultimate beneficiaries of the volumes take this opportunity to acknowledge the publisher for presenting these books that could be useful for teaching, research and extension of genetics, genomics and breeding.

Chittaranjan Kole
Preface to the Volume

Vegetable Brassicas are of economic importance worldwide. They along with cereals provide the basis of world supplies. Vegetable Brassicas contain beneficial phytochemicals—anticancerogenic compounds including minerals, vitamins, and fiber. *Brassica* species offer unique opportunities to widen our knowledge about plant growth, development and rapid phenotypic evolution. Among the *Brassica* species, *Brassica oleracea* and *B. rapa* make up most of the vegetable crops and offer numerous choices of edible forms within each species. On the basis of their origin, the greatest genetic and phenotypic variability of *B. oleracea* is observed in Europe, while Asia is the main area of diversification of vegetable *B. rapa* crops.

In the last century, the Brassicaceae family members acquired new roles as ideal model plants in basic scientific research. Furthermore, interspecific hybridization among *Brassica* spp. is relatively easy compared to other crucifer species and this character has been successfully applied to obtain synthetic amphidiploids, improve Brassica vegetable crops, study intergenomic relationships and introduce genetic variability in the genus. Thus, the genetic diversity of *Brassica* species has helped the development of many lines of research on molecular biology and plant genetics.

The phylogenetic relationship of *Arabidopsis thaliana* to *Brassica* crops offers an exceptional opportunity to benefit from the information and technology derived from the Arabidopsis Genome Project. It is already recognized, that the Arabidopsis genome sequence provides a valuable resource for identification of genes that may account for the genetic control of complex traits in Brassica. Besides, the small size of the *Brassica* genomes and their close relationship to *Arabidopsis*, provide excellent models for whole-genome sequencing. Evidently, there is potential to greatly increase our understanding of the *Brassica* biology by application of emerging genomic technologies.

This volume has been written by specialists with extensive experience in the genetics, molecular breeding and genomics of these very special and important crop plants. Chapter 1 provides basic information on vegetable Brassica crops. Breeding and genetic analysis are presented in Chapter 2. Diversity and molecular taxonomy is described in Chapter 3. Strategies of molecular linkage maps construction for vegetable Brassicas are introduced.
in Chapter 4, followed by molecular mapping of complex traits in Chapter 5. In Chapter 6, the paleopolyploidy nature and extensive cytogenetic studies of vegetable Brassicas are presented. Strategies and current status of *Brassica rapa* genome sequencing can be found in Chapter 7. Results of exploring the *Brassica* genome architecture are presented in Chapter 8. Transcriptomics and related topics are outlined in Chapter 9 and the growing importance of proteomics and metabolomics in Chapter 10. Finally, the field of bioinformatics, public databases and a corresponding network of tools is described in Chapter 11.

We hope that recent achievements and new technologies presented in this book will further support Brassica research development and improvement of vegetable Brassica crops.

Jan Sadowski
Chittaranjan Kole
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface to the Series</td>
<td>vii</td>
</tr>
<tr>
<td>Preface to the Volume</td>
<td>xiii</td>
</tr>
<tr>
<td>List of Contributors</td>
<td>xvii</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>xxiii</td>
</tr>
<tr>
<td>1. Basic Information on Vegetable Brassica Crops</td>
<td>1</td>
</tr>
<tr>
<td>Maria Elena Cartea, Margarita Lema, Marta Francisco, and Pablo Velasco</td>
<td></td>
</tr>
<tr>
<td>2. Classical Breeding and Genetic Analysis of Vegetable Brassicas</td>
<td>34</td>
</tr>
<tr>
<td>Nirala Ramchiary, Suhyoung Park, and Yong Pyo Lim</td>
<td></td>
</tr>
<tr>
<td>3. Diversity Analysis and Molecular Taxonomy of Brassica Vegetable Crops</td>
<td>81</td>
</tr>
<tr>
<td>Guusje Bonnema, Dunia Pino Del Carpio, and Jianjun Zhao</td>
<td></td>
</tr>
<tr>
<td>4. Molecular Linkage Maps: Strategies, Resources and Achievements</td>
<td>125</td>
</tr>
<tr>
<td>Danuta Babula-Skowrońska, Agata Cieśla, and Jan Sadowski</td>
<td></td>
</tr>
<tr>
<td>5. Molecular Mapping of Complex Traits</td>
<td>197</td>
</tr>
<tr>
<td>Małgorzata Kaczmarek, Matthew N. Nelson, and Wallace A. Cowling</td>
<td></td>
</tr>
<tr>
<td>6. Cytogenetic Studies in Vegetable Brassicas</td>
<td>257</td>
</tr>
<tr>
<td>Piotr A. Ziolkowski, Martin A. Lysak, and Waheeb K. Heneen</td>
<td></td>
</tr>
<tr>
<td>7. Brassica rapa Genome Sequencing Project: Strategies and Current Status</td>
<td>304</td>
</tr>
<tr>
<td>Jeong-Hwan Mun, Tie-Jin Yang, Soo-Jin Kwon, and Beom-Seok Park</td>
<td></td>
</tr>
<tr>
<td>8. Exploring the Paradoxes of the Brassica Genome Architecture</td>
<td>328</td>
</tr>
<tr>
<td>Isobel A.P. Parkin, and Stephen J. Robinson</td>
<td></td>
</tr>
<tr>
<td>9. The Decoding of Gene Functions—Transcriptomics Tools Toward Molecular Physiology and Breeding</td>
<td>349</td>
</tr>
<tr>
<td>Agnieszka Ludwików, Krzysztof Krasowski, and Leo Zeef</td>
<td></td>
</tr>
</tbody>
</table>
10. Proteomics and Metabolomics 372
   Lucyna H. Misztal, and Izabela Ibragimow

11. Role of Bioinformatics as a Tool for Vegetable Brassica Species 406
    Chris Duran, Zoran Boskovic, Jacqueline Batley, and David Edwards

Index 419
Color Plate Section 427
List of Contributors

Danuta Babula-Skowrońska
Institute of Plant Genetics, Polish Academy of Sciences, Strzeszyńska 34, 60-479 Poznań, Poland.
Email: dhab@igr.poznan.pl

Jacqueline Batley
School of Land, Crop and Food Sciences and, ARC Centre of Excellence for Integrative Legume Research, University of Queensland, Brisbane, QLD 4072, Australia.
Email: j.batley@uq.edu.au

Beom-Seok Park
Genomics Division, Department of Agricultural Biotechnology, National Academy of Agricultural Science, RDA, 150 Suin-ro, Gwonseon-gu, Suwon 441-707, Korea.
Email: pbeom@korea.go.kr

Guusje Bonnema
Laboratory of Plant Breeding, Wageningen University, 6708PB, Wageningen, The Netherlands.
Email: Guusje.Bonnema@wur.nl

Zoran Boskovic
Australian Centre for Plant Functional Genomics, School of Land, Crop and Food Sciences, Institute of Molecular Biosciences, University of Queensland, Brisbane, QLD 4072, Australia.
Email: Z.Boskovic@uq.edu.au

Dunia Pino Del Carpio
Laboratory of Plant Breeding, Wageningen University, 6708PB, Wageningen, The Netherlands.
Email: Dunia.Pino@wur.nl

María Elena Cartea
Department of Plant Genetics, Misión Biológica de Galicia, Spanish Council for Scientific Research (CSIC), P.O. Box 28, E-36080 Pontevedra, Spain.
Email: cartea@mibg.cesga.es
Agata Cieśla
Institute of Plant Genetics, Polish Academy of Sciences, Strzeszyńska 34, 60-479 Poznań, Poland.
Email: agataciesla@gmail.com

Wallace A. Cowling
School of Plant Biology and Institute of Agriculture, The University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia.
Email: wcowling@cyllene.uwa.edu.au

Chris Duran
Australian Centre for Plant Functional Genomics and School of Land, Crop and Food Sciences, University of Queensland, Brisbane, QLD 4072, Australia.
Email: c.duran@uq.edu.au

David Edwards
Australian Centre for Plant Functional Genomics and School of Land, Crop and Food Sciences, University of Queensland, Brisbane, QLD 4072, Australia.
Email: dave.edwards@uq.edu.au

Marta Francisco
Department of Plant Genetics, Misión Biológica de Galicia, Spanish Council for Scientific Research (CSIC), P.O. Box 28, E-36080 Pontevedra, Spain.
Email: mfrancisco@mbg.cesga.es

Waheeb K. Heneen
Department of Plant Breeding and Biotechnology, Swedish University of Agricultural Sciences, PO.B. 101; Alnarp 23053 Sweden.
Email: Waheeb.Heneen@ltj.slu.se

Izabela Ibragimow
Department of Biotechnology, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland.
Email: ibiza@amu.edu.pl

Małgorzata Kaczmerek
Institute of Plant Genetics, Polish Academy of Sciences, Strzeszyńska 34, 60-479 Poznań, Poland.
Email: mrum@igr.poznan.pl

Krzysztof Krasowski
Department of Biotechnology, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland.
Email: krasowsk@amu.edu.pl
Soo-Jin Kwon
Genomics Division, Department of Agricultural Biotechnology, National Academy of Agricultural Science, RDA, 150 Suin-ro, Gwonseon-gu
Suwon 441-707, Korea.
Email: sjkwon@rda.go.kr

Margarita Lema
Department of Plant Genetics, Misión Biológica de Galicia, Spanish Council for Scientific Research (CSIC), P.O. Box 28, E-36080 Pontevedra, Spain.
Email: mlema@mbg.cesga.es

Yong Pyo Lim
Department of Horticulture, Chungnam National University, Gung-Dong, Yuseong-Gu, Daejeon 305-764, Korea.
Email: yplim@cnu.ac.kr

Agnieszka Ludwików
Department of Biotechnology, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland.
Email: ludwik@amu.edu.pl

Martin A. Lysak
Department of Functional Genomics and Proteomics, Masaryk University, Kamenice 5, 625 00 Brno Czech Republic.
Email: lysak@sci.muni.cz

Lucyna H. Misztal
Department of Biotechnology, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland.
Email: lmisztal@amu.edu.pl

Jeong-Hwan Mun
Genomics Division, Department of Agricultural Biotechnology, National Academy of Agricultural Science, RDA, 150 Suin-ro, Gwonseon-gu, Suwon 441-707, Korea.
Email: munjh@rda.go.kr

Matthew N. Nelson
School of Plant Biology and Institute of Agriculture, The University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia.
Email: mnelson@plants.uwa.edu.au
Suhyoung Park
Vegetable Division, National Institute of Horticultural and Herbal Science, RDA, Suwon 441-440, Korea.
Email: psh@korea.kr

Isobel A. Parkin
Agriculture and Agri-Food Canada Saskatoon, Research Centre 107 Science Place Saskatoon, SK, S7N 0X2, Canada.
Email: isobel.parkin@agr.gc.ca

Nirala Ramchiary
Department of Horticulture, Chungnam National University, Gung-Dong, Yuseong-Gu, Daejeon 305-764, Korea.
Email: nrrdsc@gmail.com

Stephen J. Robinson
Agriculture and Agri-Food Canada Saskatoon, Research Centre 107 Science Place Saskatoon, SK, S7N 0X2, Canada.
Email: RobinsonS@agr.gc.ca

Jan Sadowski
Department of Biotechnology, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznan, Poland and Institute of Plant Genetics, Polish Academy of Sciences, Strzeszyńska 34, 60-479 Poznań, Poland.
Email: jsad@amu.edu.pl

Pablo Velasco
Department of Plant Genetics, Misión Biológica de Galicia, Spanish Council for Scientific Research (CSIC), P.O. Box 28, E-36080 Pontevedra, Spain.
Email: pvelasco@mbg.cesga.es

Tae-Jin Yang
Department of Plant Science, Plant Genomics and Breeding Institute, and Research Institute for Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul, 151-921, Korea.
Email: tjyang@snu.ac.kr

Leo Zeef
Bioinformatics Core Facility Faculty of Life Sciences, Michael Smith Building B.1083, The University of Manchester, Oxford Rd, Manchester M13 9PT, United Kingdom.
Email: Leo.Zeef@manchester.ac.uk
Jianjun Zhao
Laboratory of Plant Breeding, Wageningen University, 6708PB, Wageningen, The Netherlands.
Email: jianjun.zhao@wur.nl

Piotr A. Ziolkowski
Department of Biotechnology, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland.
Email: pzio@amu.edu.pl
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-DE</td>
<td>Two dimensional electrophoresis</td>
</tr>
<tr>
<td>2-D DIGE</td>
<td>Two-dimensional fluorescence difference gel electrophoresis</td>
</tr>
<tr>
<td>3-MSP</td>
<td>3-Methylsulfinylpropyl</td>
</tr>
<tr>
<td>4-MSB</td>
<td>4-Methylsulfinylbutyl</td>
</tr>
<tr>
<td>ABTS</td>
<td>2,2'-Azino-bis(3-ethylbenzthiazoline-6-sulphonic acid</td>
</tr>
<tr>
<td>AC</td>
<td>Affinity chromatography</td>
</tr>
<tr>
<td>ACK</td>
<td>Ancestral crucifer karyotype</td>
</tr>
<tr>
<td>ADH</td>
<td>Alcohol dehydrogenase</td>
</tr>
<tr>
<td>AdVaB</td>
<td>Adding Value to the UK Brassica Crop Science Community Consortium</td>
</tr>
<tr>
<td>AFLP</td>
<td>Amplified fragment length polymorphism</td>
</tr>
<tr>
<td>AgNOR</td>
<td>silver staining of nucleolar organizer region</td>
</tr>
<tr>
<td>APS</td>
<td>Acid phosphatase</td>
</tr>
<tr>
<td>ASTRA</td>
<td><em>Brassica</em> ASTRA database</td>
</tr>
<tr>
<td>ATH1</td>
<td>GeneChip Arabidopsis ATH1 Genome Array</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>BAC</td>
<td>Bacterial artificial chromosome</td>
</tr>
<tr>
<td>BAC-end sequences</td>
<td>BAC-end sequences</td>
</tr>
<tr>
<td>BG21</td>
<td>Biogreen21 Program</td>
</tr>
<tr>
<td>BGG</td>
<td><em>Brassica</em> Genome Gateway</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>BNK1</td>
<td><em>Brassica napus</em> kinase 1</td>
</tr>
<tr>
<td>Bras-EDB</td>
<td>European <em>Brassica</em> Database</td>
</tr>
<tr>
<td>BrGP</td>
<td><em>Brassica rapa</em> Genome Project</td>
</tr>
<tr>
<td>BrGSP</td>
<td>Multinational <em>Brassica rapa</em> Genome Sequencing Project</td>
</tr>
<tr>
<td>Ca</td>
<td>Calcium</td>
</tr>
<tr>
<td>CAGS</td>
<td>Conserved <em>Arabidopsis</em> genome sequence</td>
</tr>
<tr>
<td>CAPS</td>
<td>Cleaved amplified polymorphic sequence</td>
</tr>
<tr>
<td>CCP</td>
<td>Comparative chromosome painting</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>CDS</td>
<td>Coding sequence</td>
</tr>
<tr>
<td>CentBr</td>
<td><em>Brassica rapa</em> centromer satellite repeat sequences</td>
</tr>
<tr>
<td>CIM</td>
<td>Composite interval mapping</td>
</tr>
<tr>
<td>cM</td>
<td>CentiMorgan</td>
</tr>
<tr>
<td>CMS</td>
<td>Cytoplasmic male sterility</td>
</tr>
<tr>
<td>CO</td>
<td><em>CONSTANS</em></td>
</tr>
<tr>
<td>COT</td>
<td>Glutamateoxaloacetate transaminase</td>
</tr>
<tr>
<td>CP</td>
<td>Condensation pattern</td>
</tr>
<tr>
<td>CRBs</td>
<td>Centromeric retrotransposons of <em>Brassica</em></td>
</tr>
<tr>
<td>CrGC</td>
<td>Crucifer Genetics Cooperative</td>
</tr>
<tr>
<td>CSHL</td>
<td>Cold Spring Harbor Laboratory</td>
</tr>
<tr>
<td>CYP79B2</td>
<td>Cytochrome P450</td>
</tr>
<tr>
<td>DAPI</td>
<td>4′, 6-Diamidino-2-phenylindole</td>
</tr>
<tr>
<td>DArT</td>
<td>Diversity arrays technology</td>
</tr>
<tr>
<td>dbGaP</td>
<td>Genotype and phenotype database</td>
</tr>
<tr>
<td>DBP</td>
<td>Di-N-Butyl phthalate</td>
</tr>
<tr>
<td>DFFS</td>
<td>Diversity fixed foundation sets</td>
</tr>
<tr>
<td>DH</td>
<td>Doubled haploid</td>
</tr>
<tr>
<td>DIM</td>
<td>3,3′-Diindolylmethane</td>
</tr>
<tr>
<td>DMD</td>
<td>Digital Micromirror Device</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>EccDNA</td>
<td>Extrachromosomal circular DNA</td>
</tr>
<tr>
<td>ECP/GR</td>
<td>European Cooperative Program for Crop Genetic Resources Networks</td>
</tr>
<tr>
<td>EMS</td>
<td>Ethylmethane sulfonate</td>
</tr>
<tr>
<td>EQTL</td>
<td>Expression QTL</td>
</tr>
<tr>
<td>ESI</td>
<td>Electrospray ionization</td>
</tr>
<tr>
<td>EST</td>
<td>Expressed sequence tag</td>
</tr>
<tr>
<td>ET</td>
<td>Ethylene</td>
</tr>
<tr>
<td>FAFLP</td>
<td>Fluorescent amplified fragment length polymorphism</td>
</tr>
<tr>
<td>FAO</td>
<td>Food and Agricultural Organization</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence in situ hybridization</td>
</tr>
<tr>
<td>FLC</td>
<td><em>FLOWERING LOCUS C</em></td>
</tr>
<tr>
<td>FT</td>
<td><em>FLOWERING LOCUS T</em></td>
</tr>
<tr>
<td>GBIF</td>
<td>Global Biodiversity Information Facility</td>
</tr>
<tr>
<td>gDNA</td>
<td>Genomic DNA</td>
</tr>
<tr>
<td>GEO</td>
<td>Gene Expression Omnibus database</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>GISH</td>
<td>Genomic in situ hybridisation</td>
</tr>
<tr>
<td>GLS</td>
<td>Glucosinolates</td>
</tr>
<tr>
<td>GOT</td>
<td>Glutamateoxaloacetate transaminase</td>
</tr>
<tr>
<td>GSS</td>
<td>Genome survey sequences</td>
</tr>
<tr>
<td>HICF</td>
<td>High-information content fingerprinting</td>
</tr>
</tbody>
</table>
Abbreviations

HPLC  High-pressure liquid chromatography
HPLC/DAD  HPLC with diode array detector
HRI  Human Resources International
HTGS  High-throughput genome sequence
I3C  Indole-3-carbinol
IAA  Indole-3-acetic acid
IBPGR  International Board for Plant Genetic Resources
IG  Indole methyl glucosinolates
IMSORB  Integrated Market System for Oilseed Rape Breeding
In/Del  Insertion/deletion
ISSR  Inter simple sequence repeat
ITCs  Isothiocyanates
iTRAQ  Isobaric tag for relative and absolute quantitation
ITS  Internal transcribed spacer
JCVI  J. Craig Venter Institute
JIC  John Innes Center
Kb  kilo base
KBGP  Korea Brassica Genomics Project
LAP  Leucine aminopeptidase
LC  Liquid chromatography
LC/MS  Liquid chromatography/mass spectrometry
LD  Linkage disequilibrium
LF  LEAFY locus
LFY  LEAFY locus
LG  Linkage group
LINE  Long interspersed nuclear element
LS  Least-square
LTR  Long terminal repeats
MALDI  Matrix-assisted laser desorption ionization
MALDI-TOF  Matrix-assisted laser desorption/ionization time of flight
mass spectrometry
MAS  Marker-assisted selection
MBGB  Multinational Brassica Genome Project
MeJA/JA  Methyl jasmonate/jasmonic acid
Mg  magnesium
MIAME  Minimum information about a microarray experiment
MIM  Multiple interval mapping
MJ  Methyl jasmonate
ML  Maximum-likelihood method
MLPK  M locus protein kinase
MPSS  Massively parallel signature sequencing
MS  Mass-spectrometry
MS/MS  Tandem mass-spectrometry
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYA</td>
<td>Million years ago</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>NIL</td>
<td>Nearly isogenic line</td>
</tr>
<tr>
<td>NIHHS</td>
<td>National Institute of Horticultural &amp; Herbal Science</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NNK</td>
<td>Nicotine-derived nitrosamine ketone</td>
</tr>
<tr>
<td>NNK</td>
<td>Nicotine-derived nitrosamine ketone, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone</td>
</tr>
<tr>
<td>NOR</td>
<td>Nucleolar organizer region</td>
</tr>
<tr>
<td>NovoHMM</td>
<td>Hidden Markov Model for de novo peptide sequencing</td>
</tr>
<tr>
<td>OREGIN</td>
<td>Defra funded Oilseed Rape Genetic Improvement Network</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>PAC</td>
<td>P1-derived artificial chromosome</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PC1</td>
<td>Pack choi of group 1</td>
</tr>
<tr>
<td>PC-3</td>
<td>Prostate cancer</td>
</tr>
<tr>
<td>PCK</td>
<td>Proto-calepineae karyotype</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PCRBr</td>
<td>Peri-centromeric retrotransposons of Brassica rapa</td>
</tr>
<tr>
<td>PFGE</td>
<td>Pulsed field gel electrophoresis</td>
</tr>
<tr>
<td>PGI</td>
<td>Phosphoglucoisomerase</td>
</tr>
<tr>
<td>PGM</td>
<td>Phosphoglucomutase</td>
</tr>
<tr>
<td>PGRU</td>
<td>Plant Genetic Resources Unit</td>
</tr>
<tr>
<td>PM</td>
<td>Perfect-match probe</td>
</tr>
<tr>
<td>PMF</td>
<td>Peptide mass fingerprinting</td>
</tr>
<tr>
<td>PR1a</td>
<td>Pathogenesis-related protein 1a</td>
</tr>
<tr>
<td>PTM</td>
<td>Posttranslational modification</td>
</tr>
<tr>
<td>QPCR</td>
<td>Real-time quantitative PCR</td>
</tr>
<tr>
<td>QTL</td>
<td>Quantitative trait loci</td>
</tr>
<tr>
<td>QTN</td>
<td>Quantitative trait nucleotide</td>
</tr>
<tr>
<td>RAPD</td>
<td>Random(ly) amplified polymorphic DNA</td>
</tr>
<tr>
<td>RCBr</td>
<td>Rapid-cycling Brassica</td>
</tr>
<tr>
<td>rDNA</td>
<td>Ribosomal DNA</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction fragment length polymorphism</td>
</tr>
<tr>
<td>RIL</td>
<td>Recombinant inbred line</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive oxygen species</td>
</tr>
<tr>
<td>RP-HPLC</td>
<td>Reverse phase chromatography</td>
</tr>
<tr>
<td>RPS2</td>
<td>Resistance protein</td>
</tr>
<tr>
<td>SA</td>
<td>Salicylic acid</td>
</tr>
<tr>
<td>SAGE</td>
<td>Serial analysis of gene expression</td>
</tr>
<tr>
<td>SCAR</td>
<td>Sequence-characterized polymorphic DNA</td>
</tr>
<tr>
<td>SCX</td>
<td>Strong cation exchange</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SFP</td>
<td>Single feature polymorphism</td>
</tr>
<tr>
<td>SI</td>
<td>Self-incompatibility</td>
</tr>
<tr>
<td>SINE</td>
<td>Short interspersed nuclear element</td>
</tr>
<tr>
<td>siRNA</td>
<td>Small interfering RNA</td>
</tr>
<tr>
<td>SLG12</td>
<td>S locus glycoprotein</td>
</tr>
<tr>
<td>SNF</td>
<td>Sulforaphane</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>SRAP</td>
<td>Sequence-related amplified polymorphism</td>
</tr>
<tr>
<td>SSH</td>
<td>Suppression subtractive hybridization</td>
</tr>
<tr>
<td>SSR</td>
<td>Simple sequence repeat</td>
</tr>
<tr>
<td>STS</td>
<td>Sequence tagged site</td>
</tr>
<tr>
<td>STS-PCR</td>
<td>Sequence tagged site-polymerase chain reaction</td>
</tr>
<tr>
<td>TAC</td>
<td>Transformation-competent artificial chromosome</td>
</tr>
<tr>
<td>TAIR</td>
<td>Arabidopsis Information Resource</td>
</tr>
<tr>
<td>TE</td>
<td>Transposable element</td>
</tr>
<tr>
<td>TIGR</td>
<td>The Institute for Genomic Research</td>
</tr>
<tr>
<td>TOF</td>
<td>Time of flight</td>
</tr>
<tr>
<td>TR</td>
<td>Terminal repeat</td>
</tr>
<tr>
<td>TRIM</td>
<td>Terminal-repeat retrotransposons in miniature</td>
</tr>
<tr>
<td>TuMV</td>
<td>Turnip mosaic virus</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>UPGMA</td>
<td>Unweighted pair-group method with arithmetic mean</td>
</tr>
<tr>
<td>UPM</td>
<td>Universidad Politécnica de Madrid</td>
</tr>
<tr>
<td>US</td>
<td>United States</td>
</tr>
<tr>
<td>VIR</td>
<td>Vavilov Research Institute for Plant Industry</td>
</tr>
<tr>
<td>VRN2</td>
<td>VERNALIZATION2</td>
</tr>
<tr>
<td>WFP</td>
<td>Wisconsin fast plant</td>
</tr>
<tr>
<td>WGD</td>
<td>Whole-genome duplication</td>
</tr>
<tr>
<td>WGS</td>
<td>Whole genome shotgun</td>
</tr>
<tr>
<td>XSP30</td>
<td>Xylem sap protein</td>
</tr>
<tr>
<td>YAC</td>
<td>Yeast artificial chromosome</td>
</tr>
</tbody>
</table>
Basic Information on Vegetable Brassica Crops

María Elena Cartea,* Margarita Lema, Marta Francisco, and Pablo Velasco

ABSTRACT

The Brassica genus belongs to the Brassicaceae family and is economically the most important genus within the tribe Brassiceae, consisting of 37 different species. This genus includes a group of six interrelated species of worldwide economic importance, three diploid: Brassica nigra, Brassica oleracea, and Brassica rapa and three amphidiploid species: Brassica carinata, Brassica juncea, and Brassica napus. The genus is categorized into oilseed, forage, condiment and vegetable crops by using their buds, inflorescences, leaves, roots, seeds and stems. Brassica oleracea and B. rapa include most of the vegetable Brassica crops and they display many choices of edible forms within each species. The greatest genetic and phenotypic variability of B. oleracea is found in Europe, while Asia represents the main area of diversification of vegetable B. rapa crops. Vegetable Brassicas are of great economic importance throughout the world. Currently, Brassica crops together with cereals represent the basis of world supplies. Vegetable Brassica crops have the nutritional characteristics of other vegetable crops, as for example a low fat and protein content and a high value in vitamins, fiber and minerals. Besides, they show high quantities of phenolic compounds and have a singular kind of compounds which differentiate these crops from other vegetables; these compounds are called glucosinolates. These compounds gave Brassicas their antioxidant and anticarcinogenic properties. Finally, the remarkable morphological diversity of vegetable Brassica species and their relatives offers valuable opportunities to increase our knowledge about plant growth and development and our understanding of rapid phenotypic evolution. Moreover, the genetic diversity displayed by Brassica species has allowed the development of

*Corresponding author: ecartea@mbg.cesga.es
new crop types but also the substantial increase on the knowledge and research on molecular biology and plant genetics. In the 20th century, members of the Brassicaceae family acquired new roles as an ideal model plant for their use in basic scientific research and subsequently as educational tool.

**Keywords:** Breeding, Genetics, Germplasm, Nutritional value, Plant model, Taxonomy

### 1.1 Taxonomy

The *Brassica* genus belongs to the Brassicaceae family (previously known as Cruciferae) and economically is the most important genus within the tribe Brassiceae, consisting of 37 different species. The taxonomy of this genus is complex. Gómez-Campo (1999) presented a complete classification of the genus *Brassica* and its allied genera, indicating subgenera, sections, species and subspecies.

In that classification, Gómez-Campo subdivided the *Brassica* genus in two subgenera, i.e. *Brassica* and *Brassicaria*. Subsequently, the same author separated these two subgenera into two genera, i.e. *Brassica* and *Guenthera* (= subgenus *Brassicaria*), as they are morphologically very distinct. This separation is mainly based on the styal portion of the pistils always without seed primordia and other traits related to the stem and leaves (Gómez-Campo 2003). He suggested including the species belonging to subgenus *Brassicaria* under the generic denomination *Guenthera* Andr. in Bess. Later on, molecular studies (Warwick and Sauder 2005) confirmed that *Guenthera* should be separated from *Brassica*. Still, the main problem would involve the classification into sections within *Brassica* itself. According to the abovementioned findings, an updated classification of the *Brassica* genus is shown in Table 1-1.

The *Brassica* genus includes a group of six interrelated species of worldwide economic importance. U (1935) studied the cytology of the genus and established the relationships among the genomes of the six species. The three diploid *Brassica* species, *Brassica nigra* (L.) Koch (2n = 16), *Brassica oleracea* L. (2n = 18) and *Brassica rapa* L. (2n = 20), form the classic Triangle of U (see Fig. 6.1 in Chapter 6). In nature, these species have hybridized in different combinations to give rise to the three amphidiploid species, namely *Brassica carinata* A. Braun (2n = 4x = 34), *Brassica juncea* (L.) Czern. (2n = 4x = 36) and *Brassica napus* L. (2n = 4x = 38). Genomes of *B. rapa*, *B. nigra* and *B. oleracea* have been named A, B and C, respectively. Therefore, the resulting amphidiploids become AB, AC and BC for *B. juncea*, *B. napus* and *B. carinata*, respectively. These species have complex genomes, highly duplicated with intra- and inter-genomic conservation of Linkage blocks, which permit homoeologous recombination (Prakash and Hinata 1980).
Table 1-1 A taxonomic classification of *Brassica* genus, according to Gómez-Campo (1999, 2003).

<table>
<thead>
<tr>
<th>Section</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Brassica</strong></td>
<td></td>
</tr>
<tr>
<td><em>B. oleracea</em> L.</td>
<td></td>
</tr>
<tr>
<td><em>B. montana</em> Pourret</td>
<td></td>
</tr>
<tr>
<td><em>B. incana</em> Ten. subsp. <em>vulcanica</em></td>
<td></td>
</tr>
<tr>
<td><em>B. i. subsp. cazzac</em> (Ginz. and Teyb.) Trinajstic</td>
<td></td>
</tr>
<tr>
<td><em>B. villosa</em> Biv. subsp. <em>villosa</em></td>
<td></td>
</tr>
<tr>
<td><em>B. v. subsp. bitoniana</em> (Mazzola and Raimondo) R. and M.</td>
<td></td>
</tr>
<tr>
<td><em>B. v. subsp. drepanensis</em> (Caruel) Raimondo and Mazzola</td>
<td></td>
</tr>
<tr>
<td><em>B. v. subsp. tinei</em> (Lojac.) Raimondo and Mazzola</td>
<td></td>
</tr>
<tr>
<td><em>B. rupestris</em> Rafin subsp. <em>rupestris</em></td>
<td></td>
</tr>
<tr>
<td><em>B. r. subsp. brevisilqua</em> Raimondo and Mazzola</td>
<td></td>
</tr>
<tr>
<td><em>B. r. subsp. hispidi</em> Raimondo and Mazzola</td>
<td></td>
</tr>
<tr>
<td><em>B. macrocarpa</em> Guss.</td>
<td></td>
</tr>
<tr>
<td><em>B. insularis</em> Moris</td>
<td></td>
</tr>
<tr>
<td><em>B. cretica</em> Lam. subsp. <em>cretica</em></td>
<td></td>
</tr>
<tr>
<td><em>B. c. subsp. aequa</em> (Heldr. and Hal.) Snogerup et al.</td>
<td></td>
</tr>
<tr>
<td><em>B. c. subsp. lacoanica</em> Gustafsson and Snogerup</td>
<td></td>
</tr>
<tr>
<td><em>B. botteri</em> Vis. subsp. <em>botteri</em></td>
<td></td>
</tr>
<tr>
<td><em>B. b. subsp. mollis</em> (Vis.) Trinajstic</td>
<td></td>
</tr>
<tr>
<td><em>B. hilarionis</em> Post.</td>
<td></td>
</tr>
<tr>
<td><em>B. carinata</em> Braun</td>
<td></td>
</tr>
<tr>
<td><em>B. balearica</em> Pers.</td>
<td></td>
</tr>
<tr>
<td><strong>Rapa</strong> (Miller) Salmeen</td>
<td></td>
</tr>
<tr>
<td><em>B. rapa</em> L. subsp. <em>rapa</em></td>
<td></td>
</tr>
<tr>
<td><em>B. r. subsp. campestris</em> (L.) Clapman</td>
<td></td>
</tr>
<tr>
<td><em>B. r. subsp. chinensis</em> (L.) Hanelt</td>
<td></td>
</tr>
<tr>
<td><em>B. r. subsp. dichotoma</em> (Roxb.) Hanelt</td>
<td></td>
</tr>
<tr>
<td><em>B. r. subsp. narinosa</em> (Bailey) Hanelt</td>
<td></td>
</tr>
<tr>
<td><em>B. r. subsp. nipposinica</em> (Bailey) Hanelt</td>
<td></td>
</tr>
<tr>
<td><em>B. r. subsp. okinensis</em> (Lour.) Hanelt</td>
<td></td>
</tr>
<tr>
<td><em>B. r. subsp. tricolacularis</em> (Roxb.) Hanelt</td>
<td></td>
</tr>
<tr>
<td><em>B. napus</em> L.</td>
<td></td>
</tr>
<tr>
<td><em>B. juncea</em> (L.) Czern.</td>
<td></td>
</tr>
<tr>
<td><strong>Micropodium DC.</strong></td>
<td></td>
</tr>
<tr>
<td><em>B. fruticulosa</em> Cyr. subsp. <em>fruticulosa</em></td>
<td></td>
</tr>
<tr>
<td><em>B. f. subsp. dijafarsis</em> Blanco and Matarranz</td>
<td></td>
</tr>
<tr>
<td><em>B. f. subsp. dolichocarpa</em> Emberger and Maire</td>
<td></td>
</tr>
<tr>
<td><em>B. f. subsp. glaberrima</em> (Pomel) Maire</td>
<td></td>
</tr>
<tr>
<td><em>B. f. subsp. mauritania</em> (Cosson) Maire</td>
<td></td>
</tr>
<tr>
<td><em>B. f. subsp. numidica</em> Maire</td>
<td></td>
</tr>
<tr>
<td><em>B. f. subsp. pomeliana</em> Maire</td>
<td></td>
</tr>
<tr>
<td><em>B. f. subsp. radicata</em> (Desf.) Maire</td>
<td></td>
</tr>
<tr>
<td><em>B. nigra</em> (L.) Koch</td>
<td></td>
</tr>
<tr>
<td><em>B. cossianus</em> Boiss. and Reuter</td>
<td></td>
</tr>
<tr>
<td><em>B. spiniscens</em> Pomel</td>
<td></td>
</tr>
<tr>
<td><em>B. maurorum</em> Durieu</td>
<td></td>
</tr>
<tr>
<td><em>B. procumbens</em> (Poiret) O.E. Schulz</td>
<td></td>
</tr>
<tr>
<td><em>B. cadmea</em> O.E. Schulz</td>
<td></td>
</tr>
<tr>
<td><em>B. desertii</em> Danin and Hedge</td>
<td></td>
</tr>
</tbody>
</table>

Table 1-1 contd...
The genus is categorized into oilseed, forage, condiment and vegetable crops by using their buds, inflorescences, leaves, roots, seeds and stems. The same species can be utilized for several uses according to different forms or types. Four species, *B. oleracea*, *B. rapa*, *B. napus* and *B. juncea*, include crops, which have a horticultural use. The principal vegetable species is *B. oleracea*, which includes vegetable and forage forms, such as kale, cabbage, Broccoli, brussels sprouts, cauliﬂower and others; *B. rapa* includes vegetable forms, such as turnip, Chinese cabbage, and pak choi, along with forage and oilseed types; *B. napus* crops are mainly used as oilseed (rapeseed), although forage and vegetable types as leaf rape and “nabicol” are also included; finally, the mustard group is formed by three species, *B. carinata*, *B. nigra* and *B. juncea*, which are mainly used as a condiment because of their seeds although leaves of *B. juncea* are also consumed as vegetables in Asian countries (Nishi 1980; Table 1-2). *Brassica oleracea* and *B. rapa* include most of the vegetable Brassica crops and they display many choices of edible forms within each species. The greatest genetic and phenotypic variability of *B. oleracea* is found in Europe, while Asia represents the main area of diversification of vegetable *B. rapa* crops. They include the enlarged single apical bud of cabbage, the many axillary buds of Brussels sprouts, the arrested inflorescences and floral meristems of cauliﬂower and broccoli, the swollen stems of kohlrabi, the enlarged root of turnip and the enlarged and twisted leaves of pak choi and Chinese cabbage (Kalloo and Bergh 1993).

1.2 Economic Importance

FAO Statistics (FAOSTAT 2009) about vegetable Brassicas comprise “cabbage and other Brassicas” on the one hand and “cauliﬂower and broccoli” on the other hand, except for consumption data. It must be pointed out that FAO must include red, white and savoy cabbage, Chinese cabbage, Brussels sprouts and green kale under the name “cabbage and other Brassicas”. All these crops belong to *B. oleracea* species, with the exception of Chinese cabbage which, as mentioned above, is *B. rapa* (see Table 1-2). Probably, other *Brassica* species like *B. napus* (leaf rape) and *B. rapa* (turnip tops and turnip greens) are not included in this report, as well as *B. oleracea* crops that are cultivated and traded in local markets all around the world. With regard to consumption, Brassicas are included in the epigraph “Other
vegetables” which refers to vegetable Brassicas along with other minor vegetables. For this reason, discussion referred to consumption is based on “Other vegetables” data.

Table 1-2 Main Brassica vegetable species and crops and the plant part used for consumption.

<table>
<thead>
<tr>
<th>Species</th>
<th>Group</th>
<th>Common name</th>
<th>Organ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brassica oleracea</td>
<td>acepha</td>
<td>Kale, collards</td>
<td>Leaves</td>
</tr>
<tr>
<td></td>
<td>capitata capitata</td>
<td>Cabbage</td>
<td>Terminal leaf buds (heads)</td>
</tr>
<tr>
<td></td>
<td>capitata sabauda</td>
<td>Savoy cabbage</td>
<td>Terminal leaf buds (heads)</td>
</tr>
<tr>
<td></td>
<td>costata</td>
<td>Tronchuda cabbage</td>
<td>Loose heads</td>
</tr>
<tr>
<td></td>
<td>gennifera</td>
<td>Brussels sprouts</td>
<td>Vegetative buds</td>
</tr>
<tr>
<td></td>
<td>botrytis botrytis</td>
<td>Cauliflower</td>
<td>Inflorescences</td>
</tr>
<tr>
<td></td>
<td>botrytis italica</td>
<td>Broccoli</td>
<td>Inflorescences</td>
</tr>
<tr>
<td></td>
<td>gongylodes</td>
<td>Kohlrabi</td>
<td>Stem</td>
</tr>
<tr>
<td></td>
<td>albagbra</td>
<td>Chinese kale</td>
<td>Leaves</td>
</tr>
<tr>
<td>Brassica rapa</td>
<td>rapa</td>
<td>Turnip</td>
<td>Roots</td>
</tr>
<tr>
<td></td>
<td>rapa</td>
<td>Turnip greens</td>
<td>Leaves</td>
</tr>
<tr>
<td></td>
<td>rapa</td>
<td>Turnip tops</td>
<td>Shoots</td>
</tr>
<tr>
<td></td>
<td>chinensis</td>
<td>Pak choi, bok choi</td>
<td>Leaves</td>
</tr>
<tr>
<td></td>
<td>pekinensis</td>
<td>Chinese cabbage, pe-tsai</td>
<td>Leaves</td>
</tr>
<tr>
<td></td>
<td>parachinensis</td>
<td>Choy sum</td>
<td>Leaves</td>
</tr>
<tr>
<td></td>
<td>ravo</td>
<td>Broccoleto</td>
<td>Shoots</td>
</tr>
<tr>
<td></td>
<td>perviridis</td>
<td>Komatsuna, Tendergreen</td>
<td>Leaves</td>
</tr>
<tr>
<td>Brassica napus</td>
<td>pubularia</td>
<td>Leaf rape, nabicol</td>
<td>Leaves</td>
</tr>
<tr>
<td></td>
<td>napobrassica</td>
<td>Swede</td>
<td>Roots</td>
</tr>
<tr>
<td>Brassica juncea</td>
<td>rugosa</td>
<td>Mustard greens</td>
<td>Leaves</td>
</tr>
<tr>
<td></td>
<td>capitata</td>
<td>Head mustard</td>
<td>Heads</td>
</tr>
<tr>
<td></td>
<td>crispifolia</td>
<td>Cut leaf mustard</td>
<td>Leaves</td>
</tr>
</tbody>
</table>

1.2.1 Trade

Vegetable Brassicas are of great economic importance throughout the world. Currently, Brassica crops together with cereals represent the basis of world supplies. They provide beneficial phytochemicals as well as vitamins, minerals and fiber (more details further in this section). Probably Brassicas were domesticated in the Neolithic period thousands of years ago. Ancient Greeks, Romans, Indians and Chinese valued and used them extensively as a source of food and medicinal compounds. The evolution of Brassica crops was deeply marked by the particular regional preferences over the centuries. Consequently, human selection was focused on different parts of the plant, so much so that some of them are hardly recognizable as being members of the same species.

Import and export FAO (FAOSTAT 2009) estimated data reflect the economic importance of vegetable Brassicas around the world. In 2006,
approximately 2,289,843 t of vegetable Brassicas were exported, therefore representing a value of USD1,478 million and Spain being the main exporter (333,077 t) followed by China (292,754 t) and the United States of America (277,179 t) (FAOSTAT 2009). A mean price of USD 511 per metric ton of cabbage and other Brassicas and USD 822 per metric ton of cauliflower and broccoli were paid. Spain is by far the world’s main cauliflower and broccoli exporter, reaching 270,000 t with an elevated mean value of USD 1,135 per metric ton (mostly due to winter season greenhouse production), while the United States was the major exporter of cabbage and other Brassicas selling to other countries with a mean price of USD 860 per ton. On the other hand, the major importer of vegetable Brassicas is Canada (204,166 t) followed by Germany (201,439 t), Honk Kong special administrative region of the People’s Republic of China (199,879 t), the United States (183,013 t), the United Kingdom (176,914 t) and the Russian Federation (166,444 t). The United Kingdom was the only country where cauliflower and broccoli imports exceed that of the cabbage and other Brassicas, totalling 124,484 t with a value of almost USD 166 million.

1.2.2 Harvested Area

In 2007, vegetable Brassica were under cultivation in more than 142 countries around the world and they occupied more than 4.1 million ha (FAOSTAT 2009). Asia accounts for more than 75% of the global vegetable Brassica growing area, led by China with a total harvested area of 2.2 million ha, followed by India (511,600 ha) and the Russian Federation (171,700 ha) (Table 1-3). In the last 10 years, China has doubled the vegetable Brassica harvested area, in contrast to other Asian (Indonesia, Japan) and European (France, Italy, Poland) countries where the harvested area has shrunk.

1.2.3 Production

In recent decades, world vegetable Brassica production has been in continual expansion and has risen from 26.8 million t in 1961 to a record 88.3 million t in 2007 (Fig. 1-1). In crops, cauliflower and broccoli production has increased six times whereas that of the cabbage and other Brassicas has multiplied three-fold in the last five decades. Undoubtedly, China is the world’s number one producer of vegetable Brassicas, growing more that a half of all vegetable Brassica in 2007 (Table 1-3). India ranked next (10.3 million t), followed by the Russian Federation (4.1 million t), the Republic of Korea (3.0 million t), Japan (2.5 million t) and the United States (2.4 million t). From this production, around 88% of the vegetable Brassicas goes for fresh human consumption, while just 1.2% is processed, almost 4% is used as farm animal feed, and the rest is discarded as waste (FAOSTAT 2009).
### Table 1-3 Area harvested, production, and yield in the world and in the main vegetable Brassica-producing countries in 2007.

<table>
<thead>
<tr>
<th>Area harvested (ha)</th>
<th>Production (1,000 metric tons)</th>
<th>Yield (kg/ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cabbage and others</td>
<td>Cauliflower and broccoli</td>
</tr>
<tr>
<td>World</td>
<td>3,088,328</td>
<td>1,022,770</td>
</tr>
<tr>
<td>Africa</td>
<td>103,057</td>
<td>14,226</td>
</tr>
<tr>
<td>America</td>
<td>90,784</td>
<td>105,710</td>
</tr>
<tr>
<td>Mexico</td>
<td>6,100</td>
<td>22,000</td>
</tr>
<tr>
<td>United States of America</td>
<td>29,950</td>
<td>71,570</td>
</tr>
<tr>
<td>Asia</td>
<td>2,396,899</td>
<td>760,255</td>
</tr>
<tr>
<td>China</td>
<td>1,772,000</td>
<td>403,800</td>
</tr>
<tr>
<td>India</td>
<td>232,800</td>
<td>278,800</td>
</tr>
<tr>
<td>Indonesia</td>
<td>57,000</td>
<td>11,200</td>
</tr>
<tr>
<td>Japan</td>
<td>54,000</td>
<td>12,000</td>
</tr>
<tr>
<td>Korea, Democratic People's Republic of Korea</td>
<td>34,000</td>
<td>—</td>
</tr>
<tr>
<td>Korea, Republic of</td>
<td>48,000</td>
<td>—</td>
</tr>
<tr>
<td>Pakistan</td>
<td>4,800</td>
<td>11,400</td>
</tr>
<tr>
<td>Europe</td>
<td>496,603</td>
<td>131,179</td>
</tr>
<tr>
<td>France</td>
<td>7,000</td>
<td>27,000</td>
</tr>
<tr>
<td>Italy</td>
<td>17,753</td>
<td>18,038</td>
</tr>
<tr>
<td>Poland</td>
<td>36,400</td>
<td>15,500</td>
</tr>
<tr>
<td>Romania</td>
<td>46,000</td>
<td>1,800</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>170,000</td>
<td>1,700</td>
</tr>
<tr>
<td>Spain</td>
<td>8,250</td>
<td>24,500</td>
</tr>
<tr>
<td>Ukraine</td>
<td>64,000</td>
<td>1,700</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>9,800</td>
<td>16,600</td>
</tr>
<tr>
<td>Oceania</td>
<td>985</td>
<td>11,400</td>
</tr>
</tbody>
</table>

Source: FAOSTAT (2009).
1.2.4 Yield

Although Oceania and Africa are minor vegetable Brassica producing regions (accounting for 0.3 and 2.5%, respectively of world production in 2007), they provided the best yields. Oceania was the leader in cabbage yield and other Brassicas as well, with more than 47.1 t ha\(^{-1}\), far more than other regions, and Africa was the leader in cauliflower and broccoli yield with 20.4 t/ha. In cabbage and other Brassicas, some of the world’s best yields are recorded from the United Republic of Tanzania—on average, around 136.7 t ha\(^{-1}\), followed by El Salvador (91.2 t ha\(^{-1}\)), the Republic of Korea (62.5 t ha\(^{-1}\)) and New Zealand (58.7 t ha\(^{-1}\)). The best cauliflower and broccoli yield was in Kuwait with 45.6 t/ha, followed by New Zealand (42.9 t ha\(^{-1}\)) and the United Arab Emirates (34.8 t ha\(^{-1}\)).

1.2.5 Consumption

FAO data related to vegetable consumption around the world showed that in 2003 Asia consumed more that 70% of the world’s vegetable supply, which means around 540 million t, followed by Europe (85 million t) and America (67 million t). Led by China with 355 million t, the world’s top 10 vegetable consuming countries included India, the United States, Turkey, the Russian Federation, Japan, Egypt, the Islamic Republic of Iran, Italy and the Republic of Korea. However, Greece had the highest consumption per capita (275 kg per person per year), while China was very close (270 kg). Several other Mediterranean countries, viz. Lebanon (243 kg), Turkey (230 kg), Libyan Arab Jamahiriya (228 kg), Israel (221 kg) and Tunisia (199 kg), also figure among the greatest vegetable consumers.
Looking in depth at the vegetable consumption around the world, it can be observed that large differences exist between developed and developing countries (Fig. 1-2). Consumption in the world’s developed countries increased steadily in the last 42 years (from 85 million t in 1961 to 153 t in 2003). Meanwhile, the situation was completely different in the world’s developing countries. A rapid increase from 109 to 588 million t between 1961 and 2003 was accounted for, especially since the 1980s. Currently, vegetable consumption in the world’s developing countries is almost four times that of the developed world. This remarkable increase in consumption is clearly led by China. However, due to the huge population of the developing countries, vegetable consumption per capita is very similar in the world’s developed and developing countries, accounting for 115 and 118 kg per person and year, respectively. It is necessary to point out that the consumption per capita in developing countries increased progressively since the 1970s from 44 kg in 1972 to 118 kg in 2003. Therefore, this made the consumption in developing countries more, for the first time, than that of the developed world (Fig. 1-2).

![Figure 1-2](image-url) World series of vegetable Brassicas consumption between 1961 and 2003 and world series of vegetable Brassicas per capita consumption between 1961 and 2003.

1.3 Nutritional Quality

Nowadays, consumers are aware of the need for a constant supply of nutrients contained in plants to get optimal health benefits and there is a growing tendency to demand quality products with a higher added value. In this aspect, the popularity of Brassica crops is increasing because of their nutritional value and anticancer, antioxidant and anti-inflammatory properties. Vegetable Brassica crops have the general characteristics of...
other vegetable crops, as for example a low fat and protein content and a high value in vitamins, fiber and minerals. Besides, they show high quantities of phenolic compounds and have singular kinds of compounds which differentiate these crops from other vegetables; these compounds are called glucosinolates. Below, a short review about the most interesting nutritional compounds of vegetable Brassicas is presented. Data about vitamins, minerals as well as the nutritional components of the main vegetable Brassicas discussed below is based on the information provided by the USDA National Nutrient Database for Standard Reference (USDA 2008) (Table 1-4).

1.3.1 Vitamins

In general, vegetable Brassica crops have a high content of vitamins. If they are compared to 30 other vegetable crops (e.g., artichoke, beans, beet, carrot, sweet corn, onion, lettuce or spinach), Brassica crops have the highest level of vitamin C. With regard to vitamin E, only cilantro showed a higher level, while other crops like asparagus, soybean and spinach showed a similar level of this vitamin. Brassicas also have high levels of vitamin B-6 (exceeded by garlic, pepper and spinach), vitamin A and β-carotene (only exceeded by carrot), lutein, zeaxanthin and vitamin K. Vitamin C, E and carotenoids have the potential to prevent and treat malignant and degenerative diseases (Jahangir et al. 2009). Besides, Brassica crops show high levels of folate (similar to many other crops like eruca, bean, thistle, etc.), which is a scarce and important vitamin related to the reduced risk of vascular diseases, cancer and neural tube defects (Jahangir et al. 2009) and choline (exceeded by asparagus and similar to spinach, pea, garlic and eggplant).

1.3.2 Minerals

Brassica crops have been found to be rich in many minerals (Table 1-4). They have high levels of calcium, being higher than most of the compared vegetable crops. According to Lucarini et al. (1999), calcium present in Brassicas shows an excellent bioavailability, because of the low levels of oxalic and phytic acids, which make Brassicas a good source of calcium. They also show high levels of potassium, which is an essential mineral involved in different metabolic processes (i.e. synthesis of proteins, carbohydrate metabolism, etc.) as well as in the maintenance of human health (blood pressure, heart diseases, osteoporosis, etc.) and adequate levels for the rest of the minerals, including selenium, which is an important element from a biological, environmental and health point of view. Selenium deficiency is very common all around the world and nutritional supplements have been recommended to increase daily selenium intake (Pyrzynska 2009).
### Table 1-4 Nutritional composition in 100 g of edible portion of different vegetable Brassicas (raw material).

<table>
<thead>
<tr>
<th></th>
<th>Brassica oleracea</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Broccoli</td>
</tr>
<tr>
<td><strong>Water</strong> g</td>
<td>89</td>
</tr>
<tr>
<td><strong>Energy kcal</strong></td>
<td>34</td>
</tr>
<tr>
<td><strong>Protein</strong> g</td>
<td>2.82</td>
</tr>
<tr>
<td><strong>Total lipid</strong> g</td>
<td>0.37</td>
</tr>
<tr>
<td><strong>Ash</strong> g</td>
<td>0.87</td>
</tr>
<tr>
<td><strong>Carbohydrate</strong> g</td>
<td>6.64</td>
</tr>
<tr>
<td><strong>Dietary fiber</strong> g</td>
<td>2.6</td>
</tr>
<tr>
<td><strong>Sugars</strong> g</td>
<td>1.70</td>
</tr>
<tr>
<td><strong>Minerals</strong></td>
<td></td>
</tr>
<tr>
<td>Calcium mg</td>
<td>47</td>
</tr>
<tr>
<td>Iron mg</td>
<td>0.73</td>
</tr>
<tr>
<td>Magnesium mg</td>
<td>21</td>
</tr>
<tr>
<td>Phosphorus mg</td>
<td>66</td>
</tr>
<tr>
<td>Potassium mg</td>
<td>316</td>
</tr>
<tr>
<td>Sodium mg</td>
<td>33</td>
</tr>
<tr>
<td>Zinc mg</td>
<td>0.41</td>
</tr>
<tr>
<td>Copper mg</td>
<td>0.049</td>
</tr>
<tr>
<td>Manganese mg</td>
<td>0.210</td>
</tr>
<tr>
<td>Selenium mcg</td>
<td>2.5</td>
</tr>
<tr>
<td><strong>Vitamins</strong></td>
<td></td>
</tr>
<tr>
<td>Vitamin C mg</td>
<td>89.2</td>
</tr>
<tr>
<td>Thiamin mg</td>
<td>0.071</td>
</tr>
<tr>
<td>Riboflavin mg</td>
<td>0.117</td>
</tr>
<tr>
<td>Niacin mg</td>
<td>0.639</td>
</tr>
<tr>
<td>Pantothenic acid mg</td>
<td>0.573</td>
</tr>
<tr>
<td>Vitamin B-6 mg</td>
<td>0.175</td>
</tr>
<tr>
<td>Folate mcg</td>
<td>63</td>
</tr>
</tbody>
</table>
Table 1-4 contd...

<table>
<thead>
<tr>
<th></th>
<th>Brassica oleracea</th>
<th>Brassica rapa</th>
<th>Brassica napus</th>
<th>Brassica juncea</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Broccoli</td>
<td>Brussels sprouts</td>
<td>Cabbage</td>
<td>Red Cabbage</td>
</tr>
<tr>
<td>Choline</td>
<td>mg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-Carotene</td>
<td>mcg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitamin A</td>
<td>IU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lutein + zeaxanthin</td>
<td>mcg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitamin E</td>
<td>mg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitamin K</td>
<td>mcg</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Pe-tsai</th>
<th>Pak choi</th>
<th>Turnip greens</th>
<th>Broccoleto</th>
<th>Tendergreen</th>
<th>Turnips</th>
<th>Leaf rape</th>
<th>Swede</th>
<th>Mustard greens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy</td>
<td>kcal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total lipid</td>
<td>g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ash</td>
<td>g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carbohydrate</td>
<td>g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dietary fiber</td>
<td>g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sugars</td>
<td>g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minerals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calcium</td>
<td>mg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iron</td>
<td>mg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnesium</td>
<td>mg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphorus</td>
<td>mg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potassium</td>
<td>mg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sodium</td>
<td>mg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zinc</td>
<td>mg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copper</td>
<td>mg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manganese</td>
<td>mg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1-4 contd...
### Table 1–4 contd...

<table>
<thead>
<tr>
<th></th>
<th>Brassica rapa</th>
<th>Brassica napus</th>
<th>Brassica juncea</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pe-tsai</td>
<td>Pak choi</td>
<td>Turnip greens</td>
</tr>
<tr>
<td>Selenium</td>
<td>mcg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitamin C</td>
<td>mg</td>
<td>27.0</td>
<td>45.0</td>
</tr>
<tr>
<td>Thiamin</td>
<td>mg</td>
<td>0.040</td>
<td>0.040</td>
</tr>
<tr>
<td>Riboflavin</td>
<td>mg</td>
<td>0.050</td>
<td>0.070</td>
</tr>
<tr>
<td>Niacin</td>
<td>mg</td>
<td>0.400</td>
<td>0.500</td>
</tr>
<tr>
<td>Pantothenic acid</td>
<td>mg</td>
<td>0.105</td>
<td>0.088</td>
</tr>
<tr>
<td>Vitamin B-6</td>
<td>mg</td>
<td>0.232</td>
<td>0.194</td>
</tr>
<tr>
<td>Folate</td>
<td>mcg</td>
<td>79</td>
<td>66</td>
</tr>
<tr>
<td>Choline</td>
<td>mg</td>
<td>7.6</td>
<td>6.4</td>
</tr>
<tr>
<td>β-Carotene</td>
<td>mcg</td>
<td>190</td>
<td>2681</td>
</tr>
<tr>
<td>Vitamin A</td>
<td>IU</td>
<td>318</td>
<td>4468</td>
</tr>
<tr>
<td>Lutein + zeaxanthin</td>
<td>mcg</td>
<td>48</td>
<td>40</td>
</tr>
<tr>
<td>Vitamin B9</td>
<td>mg</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td>Vitamin K</td>
<td>mcg</td>
<td>42.9</td>
<td>45.5</td>
</tr>
</tbody>
</table>

1.3.3 Glucosinolates

In the past decade, much interest has been devoted to the positive effects of glucosinolates. Glucosinolates are a class of phytochemicals that are present in the Brassicaceae family, whose breakdown products are reported to possess cancer preventive activity, as well as anti-inflammatory, antibacterial (*Helicobacter pylori*), cardioprotection properties (Rosa 1999; Mithen et al. 2000; Fahey et al. 2001; Smith et al. 2005; Traka and Mithen 2008). Research impetus was initially generated when the breakdown products were found to be partly responsible for the characteristic flavor of Brassica vegetables (Fenwick et al. 1983). More recently, numerous studies have related the decreased risk of cancer to a diet rich in Brassica vegetables.

Glucosinolates are the major class of secondary metabolites found in Brassica crops. The molecule comprises a β-thioglucoside N-hydroxysulfate containing a side chain and a β-D-glucopyranose moiety. Glucosinolates can be grouped into three chemical classes: aliphatic, indolic and aromatic, according to whether their amino acid precursor is methionine, tryptophan or an aromatic amino acid (tyrosine or phenylalanine), respectively (Giamoustaris and Mithen 1996). The most important glucosinolates found in Brassica vegetables are methionine-derived glucosinolates (Mithen et al. 2003).

Although approximately 120 classes of glucosinolates have been identified in plants, each plant species contains up to four different glucosinolates in significant amounts (Fahey et al. 2001). Glucosinolates are not bioactive until they have been enzymatically hydrolyzed to various bioactive breakdown products by the endogenous plant enzyme called myrosinase, including isothiocyanates, nitriles, thiocyanates, epithionitriles, oxazolidine-2-thiones, and epithioalkanes (Grubb and Abel 2006). At physiological pH, isothiocyanates are the major products, whereas nitriles are formed at more acid pH (Halkier and Du 1997).

As components of feed and food, some glucosinolates degradation products have been recognized for long for their distinctive benefits to human nutrition and plant defense. This feature has led to considering Brassica foods as possible functional foods. The term “functional foods” describes foods that, if they are normal dietary constituents, can provide sufficient amounts of bioactive components that are valuable for health improvement.

Comparative studies of glucosinolate profiles indicate significant differences among cruciferous crops (VanEtten et al. 1976; Carlson et al. 1987; Kushad et al. 1999; Ciska et al. 2000). A summary of the major glucosinolates found in the main vegetable Brassicas belonging to *B. oleracea*, *B. rapa* and *B. napus* species has been recently reported by Cartea and Velasco (2008) (Table 1-5). Apart from the glucosinolate profile, large differences in the levels of both aliphatic and indole glucosinolates have been observed.
Table 1-5 Principal glucosinolates identified in *Brassica* vegetable crops.

<table>
<thead>
<tr>
<th>Crop</th>
<th>Aliphatic glucosinolates</th>
<th>Indole glucosinolates</th>
<th>Aromatic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GIB</td>
<td>PRO</td>
<td>SIN</td>
</tr>
<tr>
<td><em>Brassica oleracea</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White cabbage</td>
<td>+</td>
<td>+</td>
<td>+*</td>
</tr>
<tr>
<td>Savoy cabbage</td>
<td>+</td>
<td>+</td>
<td>+*</td>
</tr>
<tr>
<td>Red cabbage</td>
<td>+</td>
<td>+</td>
<td>+*</td>
</tr>
<tr>
<td>Kale</td>
<td>+</td>
<td>+</td>
<td>+*</td>
</tr>
<tr>
<td>Collard</td>
<td>+</td>
<td>+*</td>
<td>+*</td>
</tr>
<tr>
<td>Tronchuda cabbage</td>
<td>+</td>
<td>+</td>
<td>+*</td>
</tr>
<tr>
<td>Broccoli</td>
<td>+</td>
<td>+*</td>
<td>+</td>
</tr>
<tr>
<td>Brussels sprouts</td>
<td>+</td>
<td>+*</td>
<td>+*</td>
</tr>
<tr>
<td>Cauliflower</td>
<td>+</td>
<td>+*</td>
<td>+*</td>
</tr>
<tr>
<td>Kohlrabi</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><em>Brassica rapa</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turnip</td>
<td>+</td>
<td>+</td>
<td>+*</td>
</tr>
<tr>
<td>Turnip greens</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Turnip tops</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Chinese cabbage</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><em>Brassica napus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swede</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Leaf rape</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*Major glucosinolates found in each crop.

GIB: glucoiberin (3-methylsulfinylpropyl); PRO: progoitrin (2-hydroxy-3-butenyl); SIN: sinigrin (2-propenyl); GAL: glucoalyssin (5-methylsulphinylpentyl); GRA: glucoraphanin (4-methylsulphinylbutyl); GNA: gluconapin (3-butenyl); GBN: glucobrassicanapin (4-pentenyl); GIV: Glucoiberverin (3-methylthiopropyl); GER: glucoerucin (4-methylthiobutyl); GNL: glucobrassicin (3-indolylmethyl); GST: Gluconasturtiin (2-phenylethyl).

Adapted from Cartea and Velasco (2008).
in Brassica plants, due to the stage of development, tissue type and environmental conditions (Velasco et al. 2007).

Different epidemiological studies have indicated that diet and cancer are closely interlinked. Over the past 30 years, various studies have shown that fruits and vegetables contain natural phytochemicals such as glucosinolates that have anticarcinogenic properties (Block et al. 1992; Talalay and Zhang 1996; Talalay and Fahey 2001; Hecht 2000; Anilakumar et al. 2006). Consequently, cruciferous vegetables have been attracting interest in the last few years for their potential use in cancer chemoprevention (Rosa et al. 1997; Farnham et al. 2004; Smith et al. 2005). Results have consistently shown that the chemoprotective agents derived from this class of vegetables of the Cruciferae family have an influence on carcinogenesis during the initiation and promotion phases of cancer development. Similarly, reports from epidemiological studies and clinical trials also support this theory (Traka and Mithen 2008).

Isothiocyanates and indoles are two major groups of autolytic breakdown products of glucosinolates. Both of them exhibit protective activities against many types of cancer. In vitro and in vivo studies have reported that these compounds affect many stages of cancer development, including the induction of detoxification enzymes (Phase II enzymes) and the inhibition of activation enzymes (Phase I enzymes) (Zhang and Talalay 1994; Hecht 2000; Fahey et al. 2002; Anilakumar et al. 2006). Cartea and Velasco (2008) and Traka and Mithen (2008), have reviewed the most studied glucosinolates and glucosinolate degradation products and the anticarcinogenic mechanisms of isothiocyanates.

Apart from the medicinal value of these sulfur compounds, glucosinolates are responsible for the characteristic flavor and odor of Brassica vegetables. Glucosinolate-derived isothiocyanates produce a pungent and bitter flavor and sulfurous aroma playing a significant organoleptic role in Brassica products (Fenwick et al. 1983; Rosa et al. 1997; Padilla et al. 2007b). However, the direct relation between glucosinolate content and sensory properties is complex.

1.3.4 Phenolic Compounds

Besides glucosinolates, the contribution of Brassica vegetables to health improvement has generally been associated with their antioxidant property, which is mainly due to phenolic compounds (Jahangir et al. 2009). “Phenolic compounds” is a generic term that refers to a large number of compounds (more than 8,000) widely dispersed throughout the plant kingdom. Phenolics range from simple, low molecular-weight, single aromatic-ringed compounds to large and complex tannins and derived polyphenols (Crozier et al. 2006). They can be classified based on the number and arrangement
of their carbon atoms in flavonoids (flavonols, flavones, flavan-3-ols, anthocyanidins, flavanones, isoflavones and others) and non-flavonoids (phenolic acids, hydroxycinnamates, stilbenes and others) (Crozier et al. 2006), and are commonly found conjugated to sugars and organic acids. The most widespread and diverse group of the polyphenols are the flavonoids, which are built upon the C6–C3–C6 flavone skeleton. Phenolic compounds, especially flavonoids, possess different biological activities, but the most important ones are the antioxidant activity, the capillary protective effect and the inhibitory effects elicited in the various stages of a tumor. In many in vitro studies, phenolic compounds demonstrated a higher antioxidant activity than antioxidant vitamins and carotenoids (Podsedek 2007).

Podsedek (2007) has an extensive review on phenolic profiles in different Brassica species. Studies have mainly focused on broccoli florets and other B. oleracea crops such as cabbage and cauliflower (Nielsen et al. 1993; Price et al. 1997; Llorach et al. 2003; Vallejo et al. 2004) but, in the last years other species like B. rapa have been widely studied (Fernandes et al. 2007; Francisco et al. 2009). Broccoli is a source of flavonol and hydroxycinnamoyl derivatives. The main flavonol glycosides present in broccoli florets are quercetin and kaempferol 3-O-sophoroside. Three minor glucosides of these Aglycones were also detected, viz. isoquercitrin, kaempferol 3-O-glucoside and kaempferol diglucoside. The predominant hydroxycinnamoyl acids were identified as 1-sinapoyl-2-feruloylgentiobiose, 1,2-diferuloylgentiobiose, 1,2,2′-trisinapoylgentiobiose and neochlorogenic acid (Vallejo et al. 2003).

Nielsen et al. (1993) showed that cabbage contains a mixture of more than 20 compounds, seven of which have been identified as 3-O-sophoroside-7-O-glucosides of kaempferol and quercetin with and without further acylation with hydroxycinnamic acids. Llorach et al. (2003) found 22 compounds that were identified as being several derivatives of kaempferol, ferulic and sinapic acids. Similar compounds were found in B. rapa leaves with the presence of isorhamnetin (Fernandes et al. 2007; Francisco et al. 2009).

The content of polyphenols in vegetables, like the levels of other phytochemicals, may be influenced by various factors such as varieties, climatic conditions and cultural practices, maturity at harvest and storage conditions. In the case of phenolic compounds, sample preparation method is also very important. Phenolics in vegetables exist in both free and conjugated forms. Generally, only conjugated flavonoids are present in fresh vegetables, but aglycones may be found as a result of food processing. After hydrolysis, HPLC analysis showed that quercetin was the predominant flavonol aglycone in Brassica vegetables.

Some Brassicas provide a variety of anthocyanins (Jahangir et al. 2009). Red pigmentation of red cabbage and purple cauliflower is caused by anthocyanins. Cauliflower and red cabbage showed differences in their anthocyanin profiles: cyanidin-3,5-diglucoside was absent in cauliflower,
while it was well represented in red cabbage, together with the characteristic anthocyanin of *Brassica* genus (cyanidin-3-sophoroside-5-glucoside). The *p*-coumaryl and feruloyl esterified forms of cyanidin-3-sophoroside-5-glucoside were predominant in cauliflower, while the sinapyl ester was mostly present in red cabbage (Scalzo et al. 2008).

### 1.4 Brassica as a Model Plant in Genetics and Breeding

The remarkable morphological diversity of vegetable *Brassica* species and their relatives offers valuable opportunities to increase our knowledge about plant growth and development and our understanding of rapid phenotypic evolution. Plants of this genus are distinguished from many other plant taxa by their tendency to develop new morphological variants rapidly, as it becomes evident from the interfertility among common vegetables such as cabbage, cauliflower, broccoli, Brussels sprouts and kohlrabi. The genetic diversity displayed by *Brassica* species has allowed the development of new crop types and also a substantial increase in the knowledge and research on molecular biology and plant genetics. In the 20th century, members of the Brassicaceae family acquired new roles as an ideal model plant for their use in basic scientific research and subsequently as educational tools. The Brassicaceae family includes two excellent models, the thale cress *Arabidopsis thaliana* (L.), Heynh, and the rapid-cycling populations of the major *Brassica* species, also known as “Wisconsin Fast Plants”. On the other hand, interspecific hybridization among *Brassica* spp. is easy compared to other crucifer species and this fact has been successfully applied to develop synthetic amphidiploids, improve Brassica vegetable crops, clarify intergenomic relationships and create genetic variability in the genus.

#### 1.4.1 Rapid-cycling Brassica Populations

These populations are commonly known as Wisconsin Fast Plants (WFP) because they were developed at the University of Wisconsin-Madison by Paul Williams in 1970 ([http://www.fastplants.org](http://www.fastplants.org)). Williams and coworkers selected accessions of the six *Brassica* species of the U-triangle for a short time to flower and speed up seed maturation. Over multiple generations of recurrent selection and breeding, rapid-cycling populations of each of the six species were developed (Williams and Hill 1986). These populations could produce up to 10 generations of seed per year and their life cycle is completed in 35–45 days (*B. rapa*) to 60 days (*B. oleracea*). Over the past 30 years, rapid-cycling *Brassica* (RCBr) populations especially those of *B. rapa* and *B. oleracea* have been developed. The Wisconsin Fast Plants (WFP) Program utilizes rapid-cycling *B. rapa*, which has a life cycle as short as
that of *A. thaliana*, therefore making them easier to study. The short life cycle along with their small size and easiness to grow make them an ideal organism for teaching and for applied and basic plant biology research as genetics, molecular and cell biology, plant biochemistry and plant breeding.

RCBr populations were originally developed as a model for testing the plant/pathogen interactions. Musgrave (2000) reviewed the different applications where RCBr are excellent models, including research on plant disease, genetics, plant physiology and development, self-incompatibility and in vitro culture. Regarding this last feature, different techniques using rapid-cycling plants have been successful as protoplast, anthers, embryo culture techniques, and transformation procedures by using *Agrobacterium tumefaciens* and *A. rhizogenes*.

In addition, RCBr populations have been reported to be valuable laboratory tools for students to research into inheritance and selection along with evolution. A central idea of the WFP Program is that students learning biological concepts should have the opportunity to design and carry out their own experiments that involve living organisms. Populations of WFP are successfully used for genetic analysis of both Mendelian and complex traits and to study the genotype × environment interaction. Their use has also been reported successfully for studies on plant tissue culture in the educational situation. A review detailing the use of RCBr as a model system in education has been recently published by Himelblau et al. (2004). A particular application of RCBr has been the study of their use for developing crops to be grown in space. For instance, astronauts aboard the Mir space station have used RC *B. rapa* in microgravity experiments. Such studies have revealed a difference in shoot carbohydrate content and root zone hypoxia between plants growing in space and control plants growing on Earth (Stout et al. 2001).

### 1.4.2 *Arabidopsis*

*Arabidopsis thaliana* is a well-studied plant model worldwide. It was originally identified as an experimental tool by Laibach in Germany (cited by Dixon 2007) in the middle of the 20th century. Its short life cycle (average 60 days), the huge seed production, the estimated genome size of 125–130 Mb, which is the smallest among the higher plants known, and the low content of repetitive sequences are the characteristics which make this small plant an ideal model organism. This species has thus become the universal research tool for plant biology, including different aspects of metabolism, development, biochemistry, plant environment interactions, genetics and genomic analysis. *Arabidopsis* was the first plant for which the
whole genome was sequenced (Arabidopsis Genome Initiative 2000); this provides genes for many different types of genomic research and currently, many Arabidopsis genes have been cloned and characterized.

Species of the Brassica genus are closely related to the A. thaliana, which also belongs to the Brassicaceae family. The close relationship between the two genera, Arabidopsis and Brassica, is reflected by an average identity of exon sequences at the nucleotide level, which is estimated to be 87% (Cavell et al. 1998). The phylogenetic relationship of Arabidopsis to Brassica crops offers an exceptional opportunity to test if the information derived from the Arabidopsis Genome Project can be transferred to crop plants. The Arabidopsis genome sequence provides a valuable resource for identifying genes that may account for the genetic control of complex traits in Brassica. As an example, the results obtained from Arabidopsis are being applied to enhance a new insight into how important genes involved in the agronomy work and the nutritional aspects of Brassica species, including the genes responsible for head formation in cauliflower and broccoli (Lan and Paterson 2000) and the genes involved in glucosinolate biosynthesis (Gigolashvili et al. 2009). For details of comparative genomics between Arabidopsis thaliana and Brassica ssp. readers are suggested to see Chapter 9.

In addition, an increasing amount of information is currently available. This information provides links between Brassica genetic maps and the Arabidopsis genome. Recent studies have demonstrated the potential for comparisons among genetic and cytogenetic maps between Arabidopsis and different Brassica crops. The development of tools that facilitate the transfer of knowledge from Arabidopsis to vegetable Brassicas would help the breeding of these crops having desirable characteristics. For instance, expressed sequenced tags (ESTs) from Arabidopsis have been used as restriction fragment length polymorphism (RFLP) markers in B. oleracea, for comparison of the genomes of both species (Babula et al. 2003). Significant progress will be achieved in the coming years through the integration of candidate gene approaches in Brassica crops by using the detailed information now available for the Arabidopsis genome. Comparative mapping between the Brassica genomes and Arabidopsis is becoming a common procedure to identify candidate genes and markers for mapping studies and to expedite positional gene cloning (Qiu et al. 2009). Genome sequencing projects for B. rapa and B. oleracea are in process. This will provide an excellent opportunity to study the genome changes associated with the origin and evolution of these species in relation to A. thaliana.
1.4.3 Chromosome Evolution (Polyploidy Studies)

The Brassicaceae family includes one of the most flexible plants in terms of interspecific and intergeneric crosses. The use of the three diploid species of the U-triangle and their corresponding amphidiploids has been reported in scientific literature as an excellent model system for studying the mechanisms and the control of polyploidization and homoeologous recombination from a breeding perspective. Embryo rescue, protoplast fusion and other methods have been used to intercross the six Brassica species of the U-triangle. Due to the distribution of a series of closely related genomes and combined amphidiploids, the Brassica genus provides the opportunity to study rapid genome changes associated with polyploidy as well as chromosome evolution.

Each one of the amphidiploid species (B. napus, B. juncea and B. carinata) can be resynthesized by hybridizing the diploid species and then doubling the chromosomes (Song et al. 1993). This results in homozygous polyploid lines that should not segregate, but variation in DNA restriction fragments and phenotype has been observed among progeny derived from self-pollination of single resynthesized polyploids (Song et al. 1995). In addition, the close relationship with Arabidopsis allows comparisons from the levels of the chromosome segment up to the sequence level. It may be possible to survey large numbers of Brassica genes for changes in expression by using DNA microarrays based on Arabidopsis gene sequences. Comparisons of newly derived polyploids with their exact diploid progenitors should provide an insight into the consequences of polyploidy on plant evolution. Genome duplication or polyploidy is a widespread phenomenon throughout the plant kingdom, and provides a mechanism and context for genetic and evolutionary adaptations. Brassica species provide excellent experimental systems to understand the consequences of genome duplication. There are well studied examples of whole-genome duplication that result in allopolyploidy (combining two or more distinct genomes) or autopolyploidy (duplication of a diploid genome). New allopolyploid species such as B. napus often combine parental phenotypes with novel traits and adaptations and the fixation of hybrid vigor.

1.4.4 Tissue Culture

Vegetable Brassica species had pioneered the development of in vitro culture techniques to produce interspecific hybrids by means of protoplast fusion as well doubled haploid lines (DH) either via anther culture or microspore
culture. Microspore derived embryogenesis was first successfully completed in broccoli from isolated anthers by Keller et al. (1975). Doubled haploid technology is widely applied in vegetable Brassica crop improvement and research programs. The use of DH lines has allowed the development of tools and resources to improve quantitative traits in a more effective way. Successful anther and microspore cultures have been reported for several *B. oleracea* crops including cabbages (Kuginuki et al. 1999) and kales (Arnison and Keller 1990). Now both methods are routine in most long term breeding programs, therefore saving the necessity of working through several generations of inbreeding. Due to the strong self-incompatibility system, most vegetable Brassica crops are outbreeding species with a high degree of heterozygosity in natural populations and open-pollinated crops. For this reason, the value of microspore culture technology in these types of crops is evident. Doubled haploids have been valuable to detect quantitative trait loci (QTLs), map genes and determine their linkage relationship to other genes important in plant breeding.

1.4.5 Transgenic Technology

Vegetable Brassica crops were amongst the first to be targeted for transgenic genetic modification, for finding traits such as herbicide resistance and modification of male sterility. Genetic transformation technology has allowed the development of novel Brassica varieties by introducing transgenes from unrelated sources, something that cannot be achieved by conventional breeding methods. Conventional Brassica breeding is an intensive and time-consuming endeavor; it takes eight to ten generations to develop a new variety. In contrast, genetic transformation provides direct means to introduce a specific gene or trait into a selected genotype without adversely affecting their desirable genetic background. Moreover, most traits introduced by the gene-transfer method are dominant. Various gene transfer methods have been highly effective. The most general approach for genetic transformation is to use *A. tumefaciens* and *A. rizhogenes*. Some transgenic plants have been recovered from major Brassica vegetables and there are abundant reports of vegetable Brassica field trials by using transgenic broccoli, cabbage, cauliflower, Chinese cabbage, kale, pak choi, swede and turnip cultivars with enhanced resistance or horticultural breeding (see for review Christey and Braun 2008). While transgenic oilseed Brassica crops (such as canola) are widely grown, as far as we know, currently there are no commercial genetically-modified vegetable Brassica crops, although according to the US National Sustainable Agriculture Information Service ([www.attra.ncat.org](http://www.attra.ncat.org)) in 2006 cabbage, cauliflower, and mustard had been added to the list under development for commercial release. Notable advances in the improvement of Brassica crops should be achieved by using
Basic Information on Vegetable Brassica Crops

1.4.6 Breeding Systems and Technologies

1.4.6.1 Male Sterility

Cytoplasmic male sterility (CMS) occurs although infrequently in natural populations of \textit{Brassica} species but it provides an excellent tool to study genetic interactions between mitochondria and nucleus during flower development. CMS plants can be selected either by following sexual crosses between different species of the same family or by somatic hybridization between unrelated species. Several systems and sources of CMS have been characterized and used in \textit{Brassica} and CMS has increasingly replaced self-incompatibility in hybrid \textit{Brassica} production. A single restorer locus appears to be capable of restoring different forms of \textit{Brassica} CMS.

1.4.6.2 Self-incompatibility

As vegetable \textit{Brassica} species are largely outbreeding with a well described self-incompatibility system, they can provide valuable information concerning the distribution of genetic variation in outbreeding crops and in investigating and evaluating regeneration methodology. \textit{Brassica} species have been established as the model sporophytic pollen self-incompatibility system for research, with a well-characterized and diverse allelic series of “S-” alleles in both \textit{B. oleracea} and \textit{B. rapa} having strong to weak incompatibility reactions. The S-locus has been well characterized at the genomic and sequence level. Alleles of genes at the S-locus display considerable variability, with a very high level of non-synonymous substitution mutation. Comparative studies indicate that this variation appears to have occurred prior to the speciation of \textit{B. rapa} and \textit{B. oleracea}. Before the adoption of CMS systems, self-incompatibility was widely used in the development of F\textsubscript{1} hybrid Brassica crops.

1.4.7 Molecular Breeding

The ability to intercross morphotypes within \textit{B. oleracea} and \textit{B. rapa} has made them suitable organisms to conduct detailed genetic studies, including comparative mapping and QTL analysis. Besides, the size of the \textit{Brassica} genomes (the genomes of diploid \textit{Brassica} species are 3–5 times the size of the \textit{Arabidopsis} genome) (see Table 1-6) and their close relationship to \textit{Arabidopsis}, make them excellent for a whole-genome sequencing. Molecular markers have been widely used to assist breeding and selection procedures.
in numerous crops. During the last decades, great progress has been made in Brassica breeding due to a better understanding of the plant genome and the application of techniques for its manipulation at a molecular level. The development of DNA-based genetic markers and genetic mapping has made possible the study of structure, origin and evolution of Brassica genomes (Sadowski et al. 1996, Quiros et al. 2001). Marker development in Brassica crops has been an active subject since the early 1980s with the development of the first RFLP linkage maps for B. oleracea, B. rapa, and B. napus (reviewed by Quiros 2001). Genetic maps have been produced for all Brassica species although more effort was focused on B. napus, B. oleracea and B. rapa. In B. oleracea, several maps have been developed independently involving crosses between different crops. The chromosome maps are also being used for location of useful genes in vegetable Brassica crops. Among the genes being mapped by various laboratories, disease resistance, CMS restorers and oil seed quality are the most prominent ones. Eventually, markers linked to these genes will be used for marker-assisted selection, gene isolation, molecular characterization and manipulation. Until recently, there has been relatively little widespread integration of maps with public-domain sequence-tagged common markers. There is a constant effort to generate such integrated information and to standardize nomenclature and orientation of linkage groups. Different mapping populations of DH lines derived from specific crosses have been used to develop several published linkage maps (Quiros and Paterson 2004). For instance, B. oleracea mapping populations and the associated linkage map are publicly available at http://www.grc.warwick.ac.uk/Brassica/Brassica_mapping.html.

QTL have been identified for a wide range of morphological, physiological and crop traits in the different Brassica crop types. The ability to resolve QTL is dependent upon the access to a larger number of recombinant individuals (segregating populations, substitution or near-isogenic lines), as well as to high-density genetic and physical maps. There is a continuing requirement to provide the tools that allow researchers to resolve QTL to the level of physical maps, bacterial artificial chromosome (BAC) contigs, and candidate genes. This can be achieved through an

<table>
<thead>
<tr>
<th>Species</th>
<th>Genome</th>
<th>n</th>
<th>Genome size (Mbp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. rapa</td>
<td>A</td>
<td>10</td>
<td>500–550</td>
</tr>
<tr>
<td>B. nigra</td>
<td>B</td>
<td>8</td>
<td>470</td>
</tr>
<tr>
<td>B. oleracea</td>
<td>C</td>
<td>9</td>
<td>600–650</td>
</tr>
<tr>
<td>B. juncea</td>
<td>AB</td>
<td>18</td>
<td>1100–1500</td>
</tr>
<tr>
<td>B. napus</td>
<td>AC</td>
<td>19</td>
<td>1130–1240</td>
</tr>
<tr>
<td>B. carinata</td>
<td>BC</td>
<td>17</td>
<td>1540</td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td></td>
<td>5</td>
<td>125–130</td>
</tr>
</tbody>
</table>
integration of physical BAC contigs anchored with genetic markers onto common reference genetic linkage maps, as well as access to comparative genomic data between *Brassica* and *Arabidopsis* (Qiu et al. 2009). New methodologies, as single nucleotide polymorphism (SNP) markers, studies of allele-trait association and candidate gene approaches, have been reviewed by Quiros and Paterson (2004) and developed mainly in rapeseed. These advances will be applied in the next years to the breeding of other Brassica crops including vegetable types.

1.5 Germplasm and Gene Pools

1.5.1 Germplasm, Conservation and GeneBanks

The genetic resources available in Brassica crops are directly related to the limits of their primary, secondary and tertiary gene pools. According to Branca (2008), in *Brassica* genus, the primary gene pool of each crop is represented by its corresponding species of the U-triangle. The secondary gene pool is represented by the other species of the U-triangle, since different phylogenetic studies confirm the cytogenetic relationship among A, B, and C genomes. Finally, the tertiary gene pool gathers species and genera related to Brassica crops in 36 cytodemes that are able to change genetic materials by using special methodologies. The main cytodemes are represented by the genera *Diplotaxis*, *Eruca*, *Erucastrum*, *Sinapis*, among others. These gene pools can confer favorable alleles and useful characteristics in some cases, but they are considered to be non-adaptive in others.

*Brassica oleracea* and *B. rapa* have much more phenotypic and genetic diversity than *B. napus* and *B. juncea*. The main reason for such diversity is that all these crops have been selected over many centuries by producing divergent plant forms that are markedly different in their use and morphological, agronomical and nutritional traits. Morphological and molecular characterizations carried out in the last years have demonstrated a great genetic diversity in several *B. oleracea* crops such as cauliflower and broccoli (Hu and Quiros 1991; Branca 2008), kales (Cartea et al. 2003; Vilar et al. 2008), cabbages (Cassian and Echeverrigaray 2000; Padilla et al. 2007a) and *B. rapa* crops (Ren et al. 1995; Padilla et al. 2005; Zhao et al. 2005). In contrast, such diversity has not been found in vegetable *B. napus* (Cartea et al. 2005; Rodríguez et al. 2005; Soengas et al. 2006).

In Brassica crops, as it also happens in most crops, the pressure for more intensive agriculture has led to a progressive reduction of the natural genetic diversity within the gene pool of each crop, which is a process called “genetic erosion”. Besides, breeding programs have led to a reduction of the allelic variation in modern elite varieties. Selection within each species for adaptation to different uses and growing areas has
thus resulted in different pools of genetic variation within the cultivated germplasm (Gómez-Campo and Prakash 1999). Efforts have been made to collect and conserve the Brassica germplasm. Around 70 years ago, N.I. Vavilov proposed the use of wild forms as sources to improve crop plants and seed banks were created to maintain genetic diversity for many crops to be used used by researchers and breeders.

From a historical point of view, the N.I. Vavilov Research Institute for Plant Industry (VIR) at St. Petersburg (Russia) stands out. Its Brassica collection in 1996 comprised 6,879 accessions. The basic working collections are stored in St. Petersburg (Shashilova 1997).

Under the aegis of the European Cooperative Program for Crop Genetic Resources Networks (ECP/GR), a working group on Brassicas was established in 1991. The Centre for Genetic Resources, the Netherlands (CGN) developed the European Brassica Database (Bras-EDB) in the framework of the Brassica Working Group of the European Cooperative Program for Crop Genetic Resources Networks. The Bras-EDB contains passport data of most of the European Brassica collections (Boukema et al. 1997). At present, it contains passport data of about 19,600 accessions from 36 collections coming from 22 countries. This database can be accessed at (http://documents.plant.wur.nl/cgn/pgr/brasedb). The database has shown its value as a tool in developing core collections, in the identification of duplicates in the European collections and in defining a sound policy of safety duplication (Van Soest et al. 2004). Major updates of Bras-EDB were done in 2001 and 2005, supported financially by the European Commission by the project RESGEN CT99 109–112: “Brassica collections for broadening agricultural use, including characterizing and utilizing genetic variation in Brassica carinata for its exploitation as an oilseed crop”. This project was an important attempt to unify efforts on Brassica resources within the EU. The participant countries were France, Germany, Greece, Italy, the Netherlands, Portugal, Spain, Sweden and the United Kingdom. Besides documenting Brassica collections, other objectives of the project were to create a core collection of the four species included in the project (B. oleracea, B. rapa, B. napus and B. carinata); characterization and regeneration of collections with priority on accessions included in the core collection; evaluation of different properties with priority also in material included in the cores; rationalization and safety duplication of collections and recommendations for further collection (van Soest et al. 2004).

Other European countries that maintain collections of vegetable Brassica crops are Bulgaria, Croatia, the Czech Republic, Hungary, Poland, Russia, Switzerland and Turkey. There are also collections of cole crops in Israel, Ethiopia, South Africa, India, the Philippines and Taiwan (Swarup and Brahmi 2005).
In the US, Prof. P. H. Williams, at the University of Wisconsin, established in 1982 the Crucifer Genetics Cooperative (CrGC) to acquire, maintain and distribute seed stocks and pollen of Brassicaceae. Presently, the program provides seed stocks and information of rapid-cycling Brassicas, concretely the six species of the U-triangle. Most *Brassica* species are maintained at the Plant Genetic Resources Unit (PGRU) located at the Geneva campus (New York) of Cornell University and formed in 1986. The PGRU maintains some accessions from wild Brassicas and 1,480 accessions of *B. oleracea*.

The conservation of plant genetic resources in genebanks involves the selection of accessions to be conserved and the maintenance of these accessions for current and future users. Choices concerning both these questions need knowledge about the distribution of genetic diversity within and between accessions sampled from the gene pool, but also about the changes in variation of these samples as a result of regenerations. In outbreeding species, such as *B. oleracea* and *B. rapa*, the accessions within the core collection retain a high degree of heterozygosity and heterogeneity, which make the evaluation of quantitative traits difficult. Different research groups are currently developing “Diversity Fixed Foundation Sets” (DFFS) based on core collections to facilitate exploitation of the genetic variation in *B. napus*, *B. oleracea* and wild C genome *Brassica* spp. (Pink et al. 2008). Diversity fixed foundation sets have been defined as: “an informative set of genetically fixed lines representing a structured sampling of diversity across a gene pool” (http://www.brassica.info/resource/plants/diversity_sets.php). They are proposed to represent the majority of allelic variation within respective gene pools in a form appropriate for long term exploitation. They are similar to diversity core collections, but genetically fixed lines are being generated for each of the founder accessions. Genetically fixed lines are currently being produced at HRI by microspore culture and/or single seed descendent (where an accession is not responsive to double haploid production).

1.5.2 Wild Allied Species as Resource of Donor Genes

Wild allied species have been widely used in plant breeding programs to broaden the germplasm base in most crop species and as a source of donor elite genes (elite alleles). Brassica crops are derived from wild species, which still exist in nature and which are able to be intercrossed. The wild relatives of vegetable Brassicas offer a largely genetic resource for beneficial alleles controlling economically important quantitative traits.

The UPM (Universidad Politécnica de Madrid) created in 1966 a Plant Germplasm Bank, which has one of the most complete collections of wild crucifers in the world (1,027 taxa, 4,863 accessions), especially the Brassiceae
tribe, which has several species of economic interest including relatives of crops belonging to the genera *Brassica, Sinapis, Raphanus, Eruca*, and others. Most of the accessions were collected by Prof. Gómez-Campo and his collaborators. In 1982 the UPM Plant Germplasm Bank was designated the base-bank for Brassicaceae by IBPGR (International Board for Plant Genetic Resources, these days known as Biodiversity International). The Plant Germplasm Bank–UPM is partly included in the GBIF (Global Biodiversity Information Facility) database (www.gbif.es).

There are many wild relatives that have potential as sources for oil, condiments, genes for resistance to different diseases and pests. They also serve as sources for cytoplasmic male sterility for hybrid seed production. Gómez-Campo (1999) introduced the term “coenospecies” to refer to a cytogenetic concept of species and subspecies of the genus *Brassica* and closely related genera (e.g., *Raphanus, Sinapis, Eruca*). For instance, *Sinapis alba* is a source of genes for resistance against the blackleg pathogen (Sjödin and Glimelius 1988), *Moricandia arvensis* is a potential source of genes for improvement of photosynthetic ability and resistance to clubroot (Rawsthorne et al. 1998), and *Raphanus sativus* is a source of resistance genes against nematodes (Voss et al. 2000). *Orychophragmus violaceus* (L.) O.E. Schulz (genome OO, 2n = 24) is a wild crucifer belonging to the tribe Brassiceae. This species possesses agronomically valuable traits such as seed oil with a high content of palmitic acid, oleic acid and linoleic acid, a low percentage of linolenic acid and erucic acid (0.9%), and a high seed yield (Luo et al. 1994). Intergeneric F1 plants of the cross *B. napus* (genome AACC, 2n = 38) × *O. violaceus* have been produced by embryo rescue (Luo et al. 1994). In some cases, the opposite has also occurred, where desirable genes have been transferred from cultivated *Brassica* forms into non-*Brassica* species. Figure 1-3 shows in a schematic way the fertile somatic hybrids obtained within the Brassicaceae family between the six species from the U-triangle and other species belonging to other genera and tribes. A detailed summary of the intrageneric, intergeneric and intertribal somatic hybrids along with the traits of interest incorporated has been published by Glimelius (1999). The level of success in transferring useful genetic variation from wild sources through wide crosses depends on many factors: the extent of diversity that can be accessed to introduce useful variation, the risk to introduce deleterious traits, the possibility to use a particular valuable allele in different genetic backgrounds and the efficiency with which useful alleles can be transferred. However, most of the wild species are difficult to exploit in research programs mainly because of their extended vegetative phase or due to the difficulty in obtaining homozygous
Basic Information on Vegetable Brassica Crops

material (e.g., doubled haploid lines) by in vitro culture.

Acknowledgements

The authors would like to thank and dedicate this chapter to Prof. César Gómez-Campo, who recently passed away and whose remarkable work has been a great inspiration and an enormous help for all of us.

References


Genetics, Genomics and Breeding of Vegetable Brassicas


Basic Information on Vegetable Brassica Crops


Basic Information on Vegetable Brassica Crops


Classical Breeding and Genetic Analysis of Vegetable Brassicas

Nirala Ramchiary,¹ Suhyoung Park,² and Yong Pyo Lim†*

ABSTRACT

Brassicas provide diverse groups of vegetables to mankind in the form of leaves, flowers, stems, and roots. Apart from providing many nutrients including vitamins, minerals, and antioxidants, they are also a potential source of large amounts of dietary fiber. Among the cultivated Brassicas, *Brassica oleracea* and *Brassica rapa* have a maximum diversity of forms which are grown worldwide. Realizing the potential importance of vegetable Brassicas, breeders, since the early part of the 20th century were successful in improving quantity and quality of vegetable Brassicas using diverse morphological and genetic resources. Genetic analysis of economically important traits such as leaf, head size, plant architecture and resistance to insect pests and diseases could identify many genetic resources which could be incorporated into the breeding program. As a result, improved cultivars and hybrid varieties that give high yield, good quality, desired plant architecture and resistance/tolerance to biotic and abiotic stresses of cabbage, cauliflower, Brussels sprout, broccoli, Chinese cabbage, pakchoi etc were developed by breeders. In this chapter, we briefly described the early works by many researchers that involved classical genetic analysis and breeding which led to successful breeding of vegetable Brassicas.

Keywords: vegetable Brassicas, *Brassica rapa*, *Brassica oleracea*, breeding, genetic analysis

¹Department of Horticulture, Chungnam National University, Gung-Dong, Yuseong-Gu, Daejeon 305-764, Korea.
²Vegetable Division, National Institute of Horticultural and Herbal Science, RDA, Suwon 441-440, Korea.
*Corresponding author: yplim@cnu.ac.kr
2.1 Introduction

Brassicas are widely known for its diverse groups providing vegetables, fodder and vegetable oils to mankind. Brassica vegetables provide leaves, flowers, stems, and roots to man, which are eaten raw or cooked and as a preserved food. Today perhaps vegetable brassicas are the most important vegetable crops grown worldwide as they are a highly diversified group of crops belonging mainly to the species Brassica oleracea and B. campestris, and to a lesser extent B. juncea and B. napus. They are rich in many nutrients including vitamins, minerals, and antioxidants and largely supply dietary fiber to mankind. Among the four species of Brassicas grown for vegetables, B. oleracea and B. rapa have maximum diversity of forms, which are grown worldwide. However, predominance of vegetable brassica cultivation in the western hemisphere including Europe goes to B. oleracea, while B. rapa is mostly cultivated in Asian countries especially in China, Korea and Japan. The vegetable form of B. juncea cultivation and its evolution to different forms are mostly limited to China. The cultivated vegetable B. oleracea are found in various types, which are largely divided into six groups i.e. kales comprising green kale, marrow stem kale, collards mainly used for edible forage (var. acephala); cabbages (var. capitata, var. sabauda, var. bullata) including headed cabbages, brussel sprouts, savoy cabbage, and others; kohlrabi (var. gongylodes); inflorescence kales (var. botrytis, var. italica) including cauliflower, broccoli, sprouting broccoli, and others; branching bush kales (var. fruticosa); and Chinese kale (B. alboglabra), used as a leaf vegetable (Snogerup et al. 1990; Rakow 2004). Seven groups of vegetable B. rapa types are known, and these are: var. campestris, var. pekinensis, var. chinensis, var. parachinensis, var. narinos, var. japonica and var. rapa. Chinese cabbage and bok choy is cultivated in China, Korea and in Japan. In Korea Chinese cabbage is eaten in the form of preserved food “Kimchi” and is regarded as a national food. Turnip and turnip greens, which also belong to B. rapa, are cultivated worldwide. Different groups within B. oleracea and B. rapa are intercrossable, and the genetic study and transfer of traits are easy. In this chapter we discuss the early genetic studies and breeding of vegetable brassicas especially in different subgroups of B. oleracea and B. rapa. However, due to limited literature, only a few major crops belonging to these two species are covered. For details of the origin and diversity see Chapter 1.

2.2 Brassica oleracea

2.2.1 Early Development and Breeding History of Brassica Vegetables

Different Brassica vegetables have different origins and breeding history. It is believed that most of the present day cultivated forms of B. oleracea have originated from native wild forms, which were found growing originally
in the Atlantic coasts of Europe especially northern Spain, western France, southern and southwestern Britain (Snogerup 1980; Snogerup et al. 1990; Hodgkin 1995; Lázaro and Aguinaglade 1998). Hodgkin (1995) believed that this wild species had been brought to the Mediterranean region from the western Atlantic European coast where selection for different morphological types created the present diverse cultivated forms of *B. oleracea*. Further, it is believed that all cole crops originated from the Mediterranean regions through mutation and introgression from wild species during evolution or human selection (Swarup and Brahmi 2005; for details of origin see Chapter 1). *B. oleracea* evolved into a wide range of various subgroups or cultivars that are presently cultivated worldwide for edible roots, leaves, flower stalks or swollen stem. It is shown that plants from different subgroups belonging to this species are closer genetically than morphologically as they hybridize easily among each other. Breeding for different types, forms, quality, duration and other morphological traits by people from different regions created lots of diversity within each subgroup.

In the following section the early breeding history of a few major *B. oleracea* crops are described. The authors adapted most of the contents from the review of Shinohara (1984) for the breeding history due to limited literature from other sources/forms.

### 2.2.1.1 Cabbage

Cabbage is grown for the head formed by several wrapper leaves surrounding the terminal bud. Cabbage or commonly known as heading cabbage developed into a wide range of basic groups, types and forms in West Europe during the 12th to 16th centuries (Shinohara 1984). It is from there that different types of cabbage developed such as common cabbage, red cabbage and Savoy cabbage. The common cabbage group is the widest in diversity, with green leaves, and consists of different head forms such as pointed, drum, oval or round heads of different duration for bolting and flowering and leaf thickness. The red cabbage group is mostly round shaped, thick leaved and susceptible to hot weather, deep purplish color leaves with different duration or maturity. Savoy cabbages are with crinkled or curling green leaves and deep veins, mostly drumhead shaped.

Cabbage was first introduced to America during the 16th century from Europe and supply of seed materials from Europe continued till World War I. The breeding of brassica vegetables crops in the United States started with the basic material such as cultivars or varieties developed in Germany, Denmark or the Low countries (Shinohara 1984; Dickson and Wallace 1986). The first local variety bred was Charleston Wakefield or Large Wakefield from unfixed population of early Jersey Wakefield in the 1880s (Shinohara 1984). After World War I due to the blockade of seed supply from Europe,
the US Government research institutes and people in the seeds business collaborated to develop superior seed stocks for every vegetable crop and established suitable locations for breeding vegetables and started various activities of local vegetable breeding. As a result, the Pacific side of North-West America like Oregon and California were designated as cabbage seed production areas during that time. Breeding for cabbage yellows (Fusarium wilt) resistance started in the United States due to the outbreak of this disease during the beginning of the 20th century. Jones and Gilman in 1911 selected two plants (from Wisconsin grown Hollander and Danish Ballhead cultivars) which survived from highly infected soils at the Agricultural Experimental Station of Kensha County, and made a cross between them. F₂ plants showing 2 to 23% (as against parents 89%) infection were selected for further research by growing the plants in infested soil in 1914. Further selection was done from this basic material up to 1919, by Jones, Walker and Tisdale of Wisconsin Agricultural Experiment Station (Shinohara 1984). This first work of successful resistance breeding encouraged systematic and sophisticated resistance breeding of vegetables in the United States. The Wisconsin Hollander is recorded since 1920. According to Shinohara (1984) the nine most popular varieties cultivated in America before World War II were: Early Jersey Wakefield (early pointed head variety listed), Charleston Wakefield (medium, early, pointed head), Copenhagen market (early, round head), early Winnistadt (medium, pointed but broad based conic head), Glory of Enkhuizen (medium early, round head), All Seasons Vandergaw (medium late, deep drum head), Late Flat Dutch (derived from Brunswick (late, typical drumhead), Danish Ballhead (a strain of Amager —late, round head) and Wisconsin Hollander (late, flatted round head). It is reported that during World War II, the United States became not only self-sufficient to produce sufficient seeds for domestic markets but also became the world’s cabbage seed supplier to the allied countries replacing the West European countries.

Until World War II introduction of European, American and Japanese cabbage cultivars was reported in China and by the 1970s cultivation of local varieties were recorded (Shinohara 1984). In Japan, heading cabbage was reported to be introduced from Europe and the US in 1868. Shinohara (1984) reported that due to the climatic similarities with those of Europe and the US, Hokkaido and the north part of Tohoku were the first to grow heading cabbage and develop acclimatized local varieties such as Sapporo Daikyu (large drumhead) and Nambu (medium drumhead) from Vandergaw and Late Flat Dutch at the beginning of the 20th century. Early types of cabbage were developed from many imported cultivars such as Copenhagen Market, Early Crops, etc. However, premature bolting was a big problem in other parts of Japan due to unfavorable climatic conditions. To overcome this problem extensive effort was paid to breed
non-bolting heading cabbage. Kurato Nakano near Tokyo and Jiro Ishi from Shizuoka bred non-bolting early types. Kurato Nakano bred two varieties of NAKANO-WASE, the earliest and the early cabbage from the progenies derived from a cross between Early Spring and Sutton’s Earliest (in the year 1916) and progenies of Henderson’s All Head Early and some lines of the former (in the year 1921). Jiro Ishi bred TOYODA-WASE from progenies of hybrid between Henderson’s Early Spring and Sutton’s Flower of Spring in 1917. These two Types, NAKANO-WASE and TOYODA-WASE were the most popularly cultivated non-bolting cultivars developed indigenously in Japan for the autumn season and extended worldwide for its cultivation. These success stories encouraged others to breed more local non bolting cabbage (Shinohara 1984). The improved varieties developed by them were dark green types because these types were tolerant to diseases and easier to handle. Looking at the success of development of autumn cultivars, efforts were made to develop summer growing cultivars using Early Summer and the progenies of different crosses. As a result Nozaki-Natsumahi was developed by T. Nozaki in 1916 in Chita peninsula, Takashi at Atsumi peninsula in around 1930, and Senshu in Senshu in 1920. However, the systematic seed production and breeding of local varieties of vegetables including cabbage came into force after World War II in 1940 due to the blockade of seed supplies from other countries. The Japanese Government consolidated all Japanese seed companies of all of Japan into 18 central seed management companies and forced them to set up systematic breeding and seed testing farms with the help of central and local government research institutes. This project involved collection, analysis and systematic classification of all the vegetables grown in Japan along with their ecological history to identify superior seed resources and set ecologically suited seed production areas specific to crops. This project resulted in the systematic development of seed production in Japan, which not only made it self-sufficient but also the world’s best exporter of cabbage seeds. Development of varieties for year-round growing season was achieved between the 1950s and 1960s by selection of traits for adaptability to a particular environment or combining the traits from different genotypes and made suitable for growing in different seasons and environments. The first commercial F₁ hybrids were produced by Takii seed company in the 1940s as O-S cross. The more improved F₁ hybrid with heterosis, uniformity and wide adaptability for every season named SHIKIDOORI was developed by Takii Seed Company in 1955 that became widely adapted in the whole of Japan. After that several F₁ hybrids were developed using the local and imported seed parents from foreign countries during 1955 to 1965. This hybrid breeding became one of the best choices for breeders as in heterotic F₁ hybrids they could combine traits from both the parents, besides being
able to produce easily using self-incompatibility. Later these hybrids were exported to foreign countries as well.

2.2.1.2 Cauliflower

Cauliflower is believed to be originated from broccoli (Crisp 1982; Gray 1982). The economic part of the cauliflower is curd. The first cauliflower cultivation known as “Originals” was reported from Mediterranean Europe especially in Italy during the 15th century (Swarup and Chatterjee 1972). The spread of this crop to France and the United Kingdom took place during the 16th century with the name of “Cyprus kale”. During the 17th and 18th century cauliflower spread all over Europe including Germany and the Netherlands and many local varieties were developed from them. The notable examples are “Northern” in Yorkshire and Derbyshire, the “Cornish” in Cornwall, the “Angers” and “Roscoff” in Brittany, and “Erfurt” or its allied “Snowball” in Germany and in the Netherlands (Swarup and Chatterjee 1972). The chief characteristics of the different types of cauliflowers developed in different regions are provided in detail by Swarup and Chatterjee (1972) and by Crisp (1982) and are presented in Table 2-1 and Table 2-2. Cornish were the main favorite in the United Kingdom until 1920 when it was replaced by white curded “Roscoff” from France. In the US, commercial cultivation of cauliflower started in 1920 and both the open-pollinated cultivars and F₁s were selected from snowball cultivars derived from the Erfurt type that was characterized by small compact plants and white compact curds (Branca 2008). Mid season cultivars with larger plants were derived from the French type and late autumn season cultivars with larger head were derived from the Italian types. In India, cauliflower was introduced in 1822 from the United Kingdom and in 1824 from South Africa while in Japan, cauliflower was introduced during the late 19th century (Swarup and Chatterjee 1972). In Italy, cauliflower is believed to have been first cultivated/domesticated in the form of “Originals” that consisted of all types of variations including flower stalk masses and curd colors.

Only early to medium early types of cauliflower are grown in western European countries due to the cold climatic conditions. These are grown mainly in spring and harvested in late summer up to autumn, while around the English channels late varieties of broccoli are grown in early autumn and harvested in spring after over-wintering (Shinohara 1984). The development of medium and medium-late types of cauliflower were developed by the French in the south of France and in its colonized land of Algeria where the winter is mild enough to harvest cauliflower from late autumn through winter. Cauliflower is highly sensitive to frost during harvesting time and for seed production. So its cultivation has been popularized in subtropical Asia and the Indian subcontinent in South East Asia due to the long moderate
monsoon and a humid climate (neither too cold nor too hot) which is best suited for cauliflower cultivation. Lots of varieties such early to medium to late have been developed in these areas especially in India, China and Taiwan (Shinohara 1984; Sharma et al. 2005). In India, cauliflower is a major vegetable crop and all types of varieties ranging from early to late have been developed of which the earliest variety is “Patna Early” also known as a tropical type and it is believed that Taiwan’s earliest variety and Japan’s extra early varieties have been developed from Patna Early. T. NOZAKI bred the NOZAKI cauliflower variety from Early Snow Ball in 1916 and many varieties were available after the 1920s. The breeding of many local varieties of cauliflower was achieved by several people in the 1940’s. The notable examples are Masuda early, medium and late by T. Masuda. The development of F₁ hybrid cauliflower was started in Japan by many

Table 2-1 Characteristics of different cauliflower groups given by Swarup and Chatterjee (1972).

<table>
<thead>
<tr>
<th>Cauliflower types</th>
<th>Origin</th>
<th>Probable date of cultivation</th>
<th>Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Originals or Italians</td>
<td>Mediterranean</td>
<td>16th Century</td>
<td>Plants short; leaves erect, broad with round tips, bluish green; curds good, not protected.</td>
</tr>
<tr>
<td>Erfurt &amp; Snowball</td>
<td>Germany &amp; the Netherlands</td>
<td>18th Century</td>
<td>Plants dwarf; leaves short, erect glaucous green; curds solid, well protected.</td>
</tr>
<tr>
<td>Cornish</td>
<td>England</td>
<td>Early 19th Century</td>
<td>Plants vigorous, long stalked; leaves loosely arranged, broadly wavy; curds flat, ir regular, loose, not protected, yellow, highly, flavored.</td>
</tr>
<tr>
<td>Northern</td>
<td>England</td>
<td>19th Century</td>
<td>Leaves petiolate, broad, very wavy, serrated; curds good, well protected.</td>
</tr>
<tr>
<td>Roscoff</td>
<td>France</td>
<td>19th Century</td>
<td>Plants short; leaves long, erect, slightly wavy with pointed tip, midrib prominent, bluish green; curds white or creamy, hemispherical, well protected.</td>
</tr>
<tr>
<td>Angers</td>
<td>France</td>
<td>19th Century</td>
<td>Leaves very wavy, serrated, grayish green; curds solid, white, well protected.</td>
</tr>
<tr>
<td>Indian cauliflower</td>
<td>India</td>
<td>Late 19th Century</td>
<td>Plants short, long stalked; leaves loosely arranged, broadly wavy; curds flat, somewhat loose, yellow to creamy, not protected, highly flavored.</td>
</tr>
</tbody>
</table>

*Adapted from Sharma et al. (2005) originally given by Swarup and Chatterjee (1972).
seed companies by using introduced European Early Snow Ball cultivar (maturing in October) and tropical early cultivars from Taiwan as parents to get extra early F1 hybrids (which are harvested in mid-September) in the 1950s. Presently Japan has all types of cauliflower varieties ranging from early to late. Most of the cauliflowers are white curded but green or violet curded ones are also found among Italian landraces. At present most of the open pollinated cultivars and landraces are replaced by more uniform and high curd quality F1 hybrids.

Table 2-2 Characteristics of different cauliflower groups given by Crisp (1982).

<table>
<thead>
<tr>
<th>Group</th>
<th>Chief characteristics</th>
<th>Common types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italian</td>
<td>Very diverse, include both annuals and biennials and curds with peculiar conformations and colors</td>
<td>Jezzi, Naples (Autumn Giant), Romanesco, Flora Blanca</td>
</tr>
<tr>
<td>North-West European biennials</td>
<td>Derived within the last 300 years from Italian material</td>
<td>Old English, Walcheran, Roscoff, Angers, St. Malo</td>
</tr>
<tr>
<td>North European annuals</td>
<td>Developed in Northern Europe for at least 400 years. Origin unknown, perhaps Italian or Eastern Mediterranean</td>
<td>Lecerf, Alpha, Mechelse, Erfurt, Danish</td>
</tr>
<tr>
<td>Asian</td>
<td>Recombinants of European annuals and biennials developed within 250 years, adapted to tropics</td>
<td>Four maturity groups are recognized by Swarup and Chatterjee (1972)</td>
</tr>
<tr>
<td>Australian</td>
<td>Recombinants of European annuals and biennials and perhaps Italian stock, developed during the last 200 years</td>
<td>Not yet been categorized</td>
</tr>
</tbody>
</table>

*Adapted from Sharma et al. (2005) originally given by Crisp (1982).

2.2.1.3 Broccoli

Broccoli is grown for its economic part known as the curd. Viani (1929) reported that broccoli resulted from wild cabbage through successive improvement. The cultivation of the first form of broccoli took place along the Italian Peninsula and selection resulted in several types of sprouting and a wide range of head size (Branca 2008). Usually the curd of broccoli is green due to the the presence of chlorophyll but landraces of light red color (due to presence to anthocyanin pigment) are also found in the southernmost part of Italy. The widespread presence of landraces in this region are distinguished by harvesting time, cold requirement for flower induction, sprouting habit, leaf color and shape, head size, grain and angle of leaf curvature (Branca 2008). Cultivars with small heads with good aroma content are also found there. The presence of widespread diversity is also reported in Sicily. It is believed that from Italy and Sicily, broccoli was introduced to the United Kingdom and the northern European countries during the 17th and 18th century in Spain (Branca 2008). In North America, broccoli crops were
widespread during the 19th century, which was introduced by Italian emigrants. Most of the variety development (single terminal head type) and its cultivation have been extensively done in the US and Japan during the 1950s. Recently, its development and cultivation has been extended to South East Asia including South Taiwan, Thailand and Japan but it has been found that it has a low adaptability in subtropics as compared to cauliflower. NAKAZATO-WASE broccoli was the first local variety bred in Japan by T. NAKAZATO in 1962 (Shinohara 1984) and in the 1970s major seed companies in Japan bred many local varieties of broccoli including F₁ hybrids. Presently green head/curd F₁ hybrids with single small compact to big head having one or two secondary heads and early type are mostly cultivated. At present more uniform and high quality F₁ hybrids are quickly replacing old landraces and open-pollinated varieties developed earlier.

2.2.1.4 Brussels Sprouts

Brussels sprouts were believed to have been cultivated in Italy during the Roman times, and possibly as early as the 1200s in Belgium. The modern Brussels sprouts that we are familiar with were first cultivated in large quantities in Belgium, the capital—Brussels as early as 1587 and became prominent in the 18th century. From there, it spread to northern Europe particularly to the Netherlands and parts of UK and become established as an important vegetable crop (Dixon 2006). The development of local open-pollinated landraces took place in these areas and breeding for specific traits according to the needs were developed. Development of early maturing type and F₁ hybrids started in Japan in the 1930s, which later formed the base material for breeding cultivars suited for the freezing industry. Production of Brussels sprouts in the US began around 1800, when French settlers brought them to Louisiana. The first plantings in California’s Central Coast began in the 1920s, with significant production beginning in the 1940s. Breeding for different local types started there. With the development of the frozen food industry in the 1940s, the production of Brussels sprouts in California increased to its highest levels over the next 20 years. Presently, F₁ hybrids dominate the worldwide Brussels sprouts cultivation developed from the Netherlands, America and Japan (Dixon 2006). Current objectives of breeding is development of cultivars with regularly spaced, tight, high density sprouts, erect stem and good leaf protection.

2.2.2 Conventional Genetic Analysis in Brassica oleracea Vegetables

Although genetics was not much developed at the beginning of the 20th century, genetic analysis was done for various traits in vegetable brassicas to determine the number and the nature of gene action. The most important
traits in vegetable brassicas are shape, size, formation of head and curd, leaf traits, plant and leaf color and other plant morphologies besides diseases and pest resistances. Brassica vegetables having inter-crossable subgroups with wide morphological diversity is very suitable for genetic study by crossing them and studying the trait segregation. Taking these advantages, several genetic analyses has been done by previous researchers in *B. oleracea*, which are extensively reviewed by Ahluwalia et al. (1977), Chatterjee (1993), King (1990) and Yarnell (1956). Detailed gene lists and gene symbols of *B. oleracea* were given by Ahluwalia et al. (1977) and Yarnell (1956) while compilation of gene lists and their linkage information from morphological data was done by Wills (1977). The followings are some of the studies which showed various reports of genetic analysis in vegetable brassicas especially in *B. oleracea*. We adapted most of the references reviewed by Yarnell (1956) for genetic analysis and early linkage analysis and partly from King (1990) as many old and original research papers could not be retrieved. Genetic analysis for different traits along with the number of genes, gene action and references are also provided in Table 2-3 and Table 2-4.

### 2.2.2.1 Genetics of Head-related Traits

*B. oleracea* vegetables can be broadly classified into two groups, heading type and non-heading type. The head is an important economic part of cabbage that distinguishes it from other *B. oleracea* groups. Studies involving different genotypes for heading had different opinions about the number of genes involved in the head or heart formation in cabbage by several wrapper leaves surrounding the terminal bud. Genetic analyses by several workers have given a conclusive general agreement that heading is recessive to non-heading and intermediate condition exists in crosses between cabbage and other *B. oleracea* cultivar groups. Pease (1926) in a study involving cabbage and kale reported that heading is governed by two recessive factors/genes $n_1 n_1 n_2 n_2$. He observed F$_1$ of the cross with a rosette incurved leaves showing paint head and recovery of both the parental types in F$_2$ besides the observation of continuous variation between the two extremes. $N_1 N_1 n_2 n_2$ and $n_1 n_1 N_2 N_2$ are considered to give the same effect as $N_1 n_1 N_2 n_2$. He obtained similar results while studying cabbage and kohlrabi crosses. This was confirmed by obtaining ratios of progenies 1 $N_1 n_1 N_2 n_2$ (very slight heading):1 $N_1 n_1 N_2 n_2$ (both slight heading): 1 $n_1 n_1 N_2 n_2$ (heading) while backcrossing to cabbage. It is also reported that the loose head of the Savoy cabbage is recessive to the hard head of smooth-leaf cabbage. Allgayer (1928), in a study involving cabbage and kitchen kale cross, proposed one major dominant and three recessive minor factors for heading $K_k k_1 k_2 k_2 k_3 k_3$. In the second generation of five such crosses he got 562 heading and 168 non-heading plants indicating major dominant gene action.
<table>
<thead>
<tr>
<th>Trait</th>
<th>Crop/subspecies</th>
<th>Identified gene, if any</th>
<th>Nature of gene action</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Morphology</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plant height</td>
<td><em>B. oleracea</em></td>
<td>1 major gene</td>
<td>Tall is dominant over short</td>
<td>Pease 1926</td>
</tr>
<tr>
<td></td>
<td><em>ssp capitata</em></td>
<td>many</td>
<td>Unknown</td>
<td>Magruder 1937</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Kwan, 1934</td>
</tr>
<tr>
<td>Head shape</td>
<td><em>B. oleracea</em></td>
<td>many</td>
<td>Pointed dominant over round</td>
<td>Tschermak 1916</td>
</tr>
<tr>
<td></td>
<td><em>ssp capitata</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heading</td>
<td><em>B. oleracea</em></td>
<td>2 major gene</td>
<td>Heading is recessive to non-heading (studied in progenies derived from a cross between cabbage and kale)</td>
<td>Pease 1926</td>
</tr>
<tr>
<td></td>
<td><em>ssp capitata</em></td>
<td>4</td>
<td>1 major dominant, 3 minor recessive (studied in progenies derived from a cross between cabbage and kitchen kale)</td>
<td>Allgayer 1928</td>
</tr>
<tr>
<td></td>
<td><em>B. oleracea</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>ssp capitata</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaf width</td>
<td><em>B. oleracea</em></td>
<td>1 major gene</td>
<td>Single major dominant gene causes differences between cabbage and kohlrabi</td>
<td>Malinowski 1929</td>
</tr>
<tr>
<td></td>
<td><em>ssp capitata</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>B. oleracea</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>ssp capitata</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entire and petiolate leaf</td>
<td><em>B. oleracea</em></td>
<td>2 closely linked genes</td>
<td>Closely linked dominant genes</td>
<td>Pease 1926</td>
</tr>
<tr>
<td></td>
<td><em>ssp capitata</em></td>
<td></td>
<td>Cabbage genotype <em>petpetEnEN</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>kohlrabi</em></td>
<td></td>
<td>Kohlrabi genotype <em>PetPetEnEn</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Brussels sprouts genotype <em>PetPetEnEn</em></td>
<td></td>
</tr>
<tr>
<td>Crinkled Vs Smooth</td>
<td><em>B. oleracea</em></td>
<td>2 genes</td>
<td>Complementary gene action</td>
<td>Kwan 1934</td>
</tr>
<tr>
<td></td>
<td><em>ssp capitata</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curly and non</td>
<td><em>B. oleracea</em></td>
<td>polygenic</td>
<td>Polygenic/unknown</td>
<td>Magruder 1929</td>
</tr>
<tr>
<td>curly</td>
<td><em>ssp capitata</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>ssp acephala</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ascorbic acid</td>
<td><em>B. oleracea</em></td>
<td>several</td>
<td>Polygenic/unknown</td>
<td>Walker and Foster 1946</td>
</tr>
<tr>
<td></td>
<td><em>ssp capitata</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>B. oleracea</em></td>
<td>many</td>
<td>Early maturing dominant over late</td>
<td>Detjen and McCue 1933</td>
</tr>
<tr>
<td></td>
<td><em>ssp capitata</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait</td>
<td>Genus</td>
<td>Phenotype</td>
<td>References</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------</td>
<td>---------------------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Annual and Biennial</td>
<td>B. oleracea ssp capitata</td>
<td>many Annual is dominant over biennial</td>
<td>Detjen and McCue 1933; Walkof 1963</td>
<td></td>
</tr>
<tr>
<td>Plant colors</td>
<td>B. oleracea ssp capitata</td>
<td>4 Genotype of red cabbage is $DAbce$ and green Cabbage is $daBCE$</td>
<td>Kristofferson, 1924; Pease 1926; Malinowski 1929, Allgayer 1928</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B. oleracea ssp capitata</td>
<td>1 Dominant homozygous colored blade, heterozygous colored veins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total node number</td>
<td>B. oleracea ssp gemmifera</td>
<td>Dominant</td>
<td>Hodgkin 1981</td>
<td></td>
</tr>
<tr>
<td>Veget. tissue color (anthocyanin)</td>
<td>All B. oleracea</td>
<td>1 Dominant</td>
<td>Sampson 1967</td>
<td></td>
</tr>
<tr>
<td>Curd color</td>
<td>B. oleracea ssp botrytis</td>
<td>Recessive</td>
<td>Crisp et al. 1975</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>B. oleracea ssp capitata</td>
<td>1 recessive</td>
<td>Crisp &amp; Angell 1985</td>
<td></td>
</tr>
<tr>
<td>Orange</td>
<td>B. oleracea ssp capitata</td>
<td>&gt;1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>green</td>
<td>B. oleracea ssp capitata</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clustered flowers</td>
<td>B. oleracea ssp capitata</td>
<td>1</td>
<td>Chiang 1983</td>
<td></td>
</tr>
<tr>
<td>Rosetting</td>
<td>B. oleracea ssp capitata</td>
<td>&gt;1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-incompatibility</td>
<td>all</td>
<td>1 with more than 50 alleles Dominant recessive</td>
<td>De Nettancourt 1977</td>
<td></td>
</tr>
<tr>
<td>Stem thickness</td>
<td>B. oleracea ssp acephara</td>
<td>1 Dominant gene action</td>
<td>Allgayer 1928</td>
<td></td>
</tr>
<tr>
<td>Flower color</td>
<td>B. oleracea ssp capitata</td>
<td>1 White is simple dominant</td>
<td>Pearson , 1929</td>
<td></td>
</tr>
<tr>
<td>Clubroot</td>
<td>B. oleracea ssp capitata</td>
<td>&gt;1 Recessive and dominant</td>
<td>Weisaeth 1977; Walker &amp; Larsen 1980; Natti et al. 1967</td>
<td></td>
</tr>
<tr>
<td>Alternaria brassicae</td>
<td>B. oleracea ssp gemmifera</td>
<td>polygenic Polygenic/ unknown</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2-3 contd...
<table>
<thead>
<tr>
<th>Trait</th>
<th>Crop/subspecies</th>
<th>Identified gene, if any</th>
<th>Nature of gene action</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peronospora parasitica</td>
<td>B. oleracea ssp italic</td>
<td>1</td>
<td>Recessive and dominance</td>
<td>Hoser-Krauze et al. 1987</td>
</tr>
<tr>
<td>Turnip Mosaic Virus</td>
<td>B. oleracea ssp capitata</td>
<td>≥ 4</td>
<td>Recessive and dominance</td>
<td>Pink &amp; Walkey 1986</td>
</tr>
<tr>
<td></td>
<td>B. oleracea ssp gemmifera</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cauliflower Mosaic Virus</td>
<td>B. oleracea ssp gemmifera</td>
<td>≥ 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B. oleracea ssp gemmifera</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pest resistance</td>
<td>Delia radicum</td>
<td>&gt; 1</td>
<td></td>
<td>Ellis &amp; Hardman 1975</td>
</tr>
<tr>
<td></td>
<td>B. oleracea ssp capitata</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aphids</td>
<td>B. oleracea ssp capitata</td>
<td>&gt; 1</td>
<td></td>
<td>Ellis &amp; Hardman 1988</td>
</tr>
<tr>
<td></td>
<td>B. oleracea ssp italic</td>
<td></td>
<td></td>
<td>Ellis &amp; Hardman 1988</td>
</tr>
<tr>
<td>Lepidoptera</td>
<td>B. oleracea ssp capitata</td>
<td></td>
<td></td>
<td>Ellis &amp; Hardman 1988</td>
</tr>
</tbody>
</table>

*Prepared from Yarnell (1956) and King (1990).*
He suggested that $KK_1K_2K_3$ gives a very loose head and the recessive factors have cumulative effect for head tightness. Three genes governing the head formation was suggested by Malinowski (1929) and Samson (1930) while Kristofferson (1924) suggested many factors governing the heading trait. The variations in the proposed number of factors for heading might be due to the selection of parental genotypes in their study.

The shape of the cabbage head varies from pointed, flat, round or oblate to intermediate shapes. The genetic analysis showed that many factors are involved in governing the head shape (Detjen and McCue 1933; Magruder 1937). Tschermak (1916) reported that the pointed shape is dominant over the round one. Pearson (1934) reported that the number of non-wrapper leaves below the proper head varies from a few to as many as 25. The number of head leaves is also confounded by whether the cabbage is early

Table 2-4: Genetics analysis of some cauliflower traits.

<table>
<thead>
<tr>
<th>Character</th>
<th>Nature of gene action</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curd weight</td>
<td>Dominance and epistasis</td>
<td>Swarup and Pal 1966</td>
</tr>
<tr>
<td></td>
<td>Pronounced over-dominance and epistasis</td>
<td>Singh et al. 1975</td>
</tr>
<tr>
<td></td>
<td>Additive and dominance</td>
<td>Singh et al. 1976; Jyoti and Vashistha 1986; Gangopadhyay et al. 1997; Sharma et al. 1988</td>
</tr>
<tr>
<td>Curd to plant</td>
<td>Partial dominance</td>
<td>Kale et al. 1979</td>
</tr>
<tr>
<td>Curd diameter</td>
<td>Predominance of dominance</td>
<td>Lal et al. 1979</td>
</tr>
<tr>
<td>Curd size index</td>
<td>Pronounced over-dominance and epistasis</td>
<td>Singh et al. 1975</td>
</tr>
<tr>
<td></td>
<td>Dominance and epistasis</td>
<td>Swarup and Pal 1966; Lal et al. 1979</td>
</tr>
<tr>
<td></td>
<td>Additive dominant gene action</td>
<td>Singh et al. 1976; Sharma et al. 1988</td>
</tr>
<tr>
<td></td>
<td>Partial dominance</td>
<td>Kale et al. 1979</td>
</tr>
<tr>
<td>Curd angle</td>
<td>Pronounced additive</td>
<td>Lal et al. 1979</td>
</tr>
<tr>
<td></td>
<td>Additive and dominance</td>
<td>Dadlani 1977; Chand 1980</td>
</tr>
<tr>
<td>Curd compactness</td>
<td>Polygenic</td>
<td>Nieuwhof and Garretson 1961</td>
</tr>
<tr>
<td></td>
<td>Dominance and additive</td>
<td>Lal et al. 1979</td>
</tr>
<tr>
<td></td>
<td>Additive</td>
<td>Vashistha et al. 1985</td>
</tr>
<tr>
<td>Early maturity</td>
<td>Partially dominant gene action</td>
<td>Watts 1964</td>
</tr>
<tr>
<td></td>
<td>Dominance and epistasis</td>
<td>Swarup and Pal 1966</td>
</tr>
<tr>
<td></td>
<td>Predominance of additive</td>
<td>Singh et al. 1975; Lal et al. 1979; Mahajan et al. 1996; Gangopadhyay et al., 1997</td>
</tr>
<tr>
<td></td>
<td>Additive gene action</td>
<td>Kale et al. 1979</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sandhu and Singh 1977;</td>
</tr>
<tr>
<td>Late maturity</td>
<td>Recessive polygenes</td>
<td>Watts 1963</td>
</tr>
</tbody>
</table>

*Adapted from Sharma et al. (2005) originally given by Chatterjee (1993).
or late type, the early type having less head leaves. Crosses involving parents with a few and many non-wrapper leaves showed few wrapper leaves to be dominant in F₁, and many modifying factors governing this trait have also been reported. Head size is inherited quantitatively and crosses involving diverse or unrelated lines have been reported to increase the head size (Kotowski 1926; Pearson 1934). The size of the cabbage head is also reported to depend on day-length, long day-length giving a large head while a short day producing a small head.

The genetics of the cabbage core wide and length have also been studied. It has been found that a wide core is dominant over a narrow core. A narrow core is desirable. Core length is controlled by two independent dominant genes for a short core, a short core having < 25% of the head diameter is desirable (Dixon 2006). Another problem in cabbage is that of the head splitting at maturity, this has been reported to be governed by three genes acting additively with partial dominance for early splitting. The narrow sense heritability for splitting was found to be 47% (Chiang 1972). Observation of head splitting revealed that the long core head split at the top while the short core split at the base (Dixon 2006).

Several studies reported that annual habit is dominant over biennial (Detjen and McCue 1933; Walkof 1963). Dominance series of genes/alleles has been suggested to account for the genetic variability in vernalization requirement and earliness (Dickson and Wallace 1986). Vegetative maturity of cabbage head showed that the F₁ derived from a cross involving parents with early maturity and late maturity was early (Rasmusson 1932).

2.2.2.2 Genetics of Curd-related Traits

Curd and its related traits are important in cauliflower. Many studies on curd and its related traits were reviewed by King (1990). The curd formation trait in cauliflower may be either dominant or recessive and is reported to be governed by major genes (Crisp 1982). Polygenic inheritance of curding has also been opined from observation of non-distinct Mendelian segregation ratios in F₂s derived from crosses involving cauliflower and broccoli. Furthermore, poor curd formation in F₁s and F₂s derived from wide crosses within cauliflower lines were also observed (Gray and Crisp 1979). Involvement of some modifier genes in curd formation were also reported (Crisp 1982). Many genes governing curd color have been identified (King 1990). A dominant gene for white color over yellow has been identified. Another gene with partial dominance for white curds has also been reported that causes dwarf orange curd in double recessive condition (Crisp et al. 1975). Crisp and Angell (1985) reported an independent dominant gene for greenless curd. Singh et al. (1978) identified two genes for leaf cover of the curd affecting discoloration by sunlight, while Dickson and Lee (1980)
reported two/three dominant genes controlling persistence of whiteness in curd exposed to sunlight. The degree of “sprouting” in broccoli is under the control of many genes affecting several axillary branching (King et al. 1990).

2.2.2.3 Genetics of Plant Morphology

Influence of single major gene to several genes has been reported for plant height. Kwan (1934) reported that several genes govern plant height, while Pease (1926) reported a major dominant gene \( T \) for height. Most cultivars are recessive for this trait although some cultivars may have the dominant gene. A long stem is undesirable, but for machine harvesting an upright head and not too short a height is required (Dixon 2006). Total node number of Brussels sprouts is the result of the dominant gene action and closely correlated to the rate of node production, and the number of harvested sprouts (Hodgkin 1981).

2.2.2.4 Genetics of Leaf Morphology

A single major dominant gene that is different between cabbage and kohlrabi has been reported for leaf width (Pease 1926; Malinowski 1929). Cabbage has the dominant gene \( W \) for wide leaf. The presence of modifying factors has also been reported for this trait. The sessile entire leaf of cabbage and petiolate lyrate leaf of mature kohlrabi has been reported to be governed by two genes: \( \text{Pet/P} \) producing petiolate and \( \text{En/E} \) producing entire leaf. The \( F_1 \) (\( \text{Enen/Petpet} \)) derived from a cross between cabbage (\( \text{EnEn/petpet} \)) and kohlrabi (\( \text{enen/petPet} \)) produces entire petiolate leaves like Brussels sprouts (\( \text{EnEn/PetPet} \)). Tschermak (1916) reported that crinkled or rugose (\( W \)) leaf of Savoy cabbage is dominant over a smooth (\( S \)) leaf. Kwan (1934) suggested complementary interaction between the genes \( W \) and \( S \) in addition to the presence of modifying factors. The phenotype of double recessive for \( W \) and \( S \) is smooth. Malinowski (1929) observed continuous variation of leaf from curliness to smoothness in the second generation and suggested that the non-curly leaf of cabbage and the curly leaf of kale are polygenically controlled. Pease (1926) also found similar results. He further suggested that the presence of bizarre protuberances (also known as Asparagodes) was due to a dominant or partially dominant gene, \( A/As \).

2.2.2.5 Genetics of Plant and Flower Traits

Genetic studies of anthocyanin pigmentation by several workers in red cabbage revealed inheritance of this trait to be governed by one gene (Allgayer 1928) to many genes (Kristofferson 1924; Pease 1926; Malinowski
Allgayer (1928) obtained second generation segregation of 235 colored blade; 358 colored midrib; 166 green, is the ratio approximately equal to the monogenic segregation of 1:2:1. Two factors, $A$ and $D$, produce red cabbage, while $A$ is the basic color factor with no effect but producing color with $D$ (Kristofferson 1924; Pease 1926; Malinowski 1929). The gene $B$ alone produces red midrib, and with $A$ produces a dark violet midrib. The factor $C$ alone is colorless but with $A$ gives a dark violet color. Gene $E$ extends the area of dark violet color. $B$ and $C$ have the same effect as $B$ alone (Kristofferson 1924). Based on that, the genotypes of red cabbage and green cabbage are deduced to be $DAbce$ and $DaBCE$, respectively. An $F_1$ plant of green and dark red produces a pink color. The gene $M$ produces magenta and $S$ produces sun color (Magruder 1937).

Inheritance of sun color was shown to be governed by two duplicate genes $R_1$ and $R_2$ (Kwan 1934). In another cross involving deep purple and green, Kwan (1934) observed 9 purple; 3 sun color; 4 green in the second generation. He proposed the genotypes $GH$, $Gh$, and $gH$ and $gh$ for deep purple, sun, and green colors, respectively. Inheritance of duplicate genes for color was further confirmed by Kristofferson (1921) when he observed 9 dark reddish violets; 3 clear red, 4 white from the cross involving parents of white cabbage and kale. White flower in Cole crops is governed by a single dominant gene $Wh$ (Pearson 1929). Segregation of 1 white: 1 yellow was observed in the first generation when a normal plant with a yellow flower was crossed with a mutant with a white flower (Kakizaki 1930).

Male sterility in Cole crops has been found to result from recessive mutation ($ms$) of male fertility gene $Ms$ (Neiuwhof 1961, 1968; Borchers 1966; Ahluwalia et al. 1977). Male sterility under the control of duplicate dominant genes with cumulative effect has also been reported (Van der Meer 1985). Ruffio-Chable et al. (1993) reported dominant male sterility in cauliflower.

Self-incompatibility is a mechanism where pollen cannot fertilize an ovary in the same plant. Self-incompatibility may be gametophytic and sporophytic. In gametophytic self-incompatibility, the incompatibility reaction is determined by the genotype of the gamete while in sporophytic, incompatibility reaction is determined by the genotype of the gamete-producing plants. Genetic studies in different $B. oleracea$ vegetables showed that self-incompatibility mechanism is a complex system, which is governed by a single major gene $S$ with 50–70 alleles (Yarnell 1956; Thompson and Taylor 1966; De Nattancourt 1977; Dickson and Wallace 1986; Ordas and Cartea 2008). An incompatible reaction occurs when identical alleles are inherited by both the pollen and stigma. Recently molecular evidences showed that recognition of self is mediated by $S$-locus specific glycoproteins ($SLGs$). At present many genes involved in self-incompatibility mechanism have been cloned and characterized (Nasrallah et al. 1987, 1988; Trick
and Flavell 1989). Sporophytic self-incompatibility is widely used for F₁ hybrid vegetables production in Brassicas (Shinohara 1984; Dickson and Wallace 1986).

2.2.2.6 Genetics of Insect-Pest and Disease Resistance

Insect pests and diseases are the major factors causing crop losses not only in vegetables but also in all other crops. Breeding resistance to diseases and insect-pests involves identification of resistance sources/genotypes and transfer of resistance genes from resistant genotypes to susceptibles. Some important classical genetic analysis for identification of resistant sources for various diseases are listed below. Downy mildew caused by *Peronospora parasitica* is a soil-borne disease and this disease develops first on lower leaves as a sparse to densely packed white mat forming chlorotic to partially chlorotic lesions when the soil is transferred by wind and rain. Genetics of resistance for this disease has been found to be complicated due to the existence of several pathogenic races even though its resistance has been identified in cabbage and broccoli (Natti et al. 1967). The cabbage genotype, PI 245015, has two independent dominant genes for resistance to the races 1 and 2. Williams et al. (1972) identified the resistance source for black rot, a seed-borne disease caused by *Xanthomonas campestris*. They reported that a major gene *f* and two modifier genes, one dominant and the other recessive, govern the resistance to this disease. Powdery mildew caused by *Erysiphe polygoni* D.C. is another disease characterized by a fine necrotic flecking on the exposed head and lower leaves of cabbage. Resistance to this disease has been identified in the cabbage cultivar Globbelle, which is governed by a single dominant gene, although influence of some modifying factors was also assumed (Walker and Williams 1965).

Cabbage yellows caused by soil-borne fungus *Fusarium oxysorum* f. sp. *conglutinens* is characterized by progressive yellowing followed by brown necrosis of the lower leaves, stunted growth and premature leaf drop. Walker (1930) designated two types of resistance, Type A and Type B. Type A resistance is controlled by a single dominant gene and carried by most resistance cultivars, and type B resistance is controlled by several genes. Type A is not influenced by temperature, whereas type B resistance breaks down at 22°C. Major gene resistance was observed in All Head Early, Copenhagen Market and Glory of Enkhuizen. However, polygenic resistance was observed in Wisconsin Hollander. Resistance to stalk rot caused by *Sclerotinia sclerotiorum* in Indian cauliflower was studied in populations from six generations of six crosses and was found to be polygenic and under the control of recessive genes primarily with additive gene action (Baswana et al. 1991).
Turnip mosaic virus (TuMV) and cauliflower mosaic virus are the two major viral diseases of vegetable brassicas including cabbage, which are characterized by mottling and necrotic spots or ring spots. Complete resistance has been identified for all four races of turnip mosaic virus (identified so far) in Chinese cabbage (Provvidenti 1980). Cabbage cultivar Globelle exhibit quantitative resistance to the races 1 and 2. Non-strain-specific recessive resistance controlled by at least one gene has been reported in F1 hybrids and open-pollinated cultivars of cabbage with high heritability (Pink and Walkey 1986). The resistance to TuMV is found to be controlled by at least four genes in Brussels sprouts with high heritability (Pink and Walkey 1986). In Brussels sprouts, resistance to cauliflower mosaic virus is generally dominant and is controlled by at least four genes (Pink and Walkey 1986). However, presence of some recessive genes has also been reported.

Clubroot disease, caused by *Plasmodiophora brassicae*, is a major disease causing severe problems in both quality and quantity of *B. oleracea* crops including cauliflower, broccoli, kale and others (for detailed review of genetics of clubroot resistance in *Brassica* species see Piao et al. 2009). Different research groups identified various sources of clubroot resistance through screening of germplasm (Crute et al. 1980; Crisp et al. 1989; Dias et al. 1993; Voorrips and Kanne 1996; Manzanares-Dauleux et al. 2000; Carlsson et al. 2004). In contrast to *B. rapa*, completely resistant accessions have rarely been identified in *B. oleracea* although large numbers of accessions were screened. Crisp et al. (1989) evaluated about 1,000 *B. oleracea* accessions and confirmed resistant sources in some kales including the curly, marrowstem and 1000-head types and cabbage including Böhmerwaldkohl, Bindsachsener cabbage, Badger Shipper and Ladoszkaya cabbage. Also some open-pollinated Brussels sprouts including Cambridge, Continuity, Rubine and Catskill and forms of South European cabbage, cauliflower and broccoli exhibited lower levels of susceptibility. Among kale, cabbage and winter cauliflower accessions evaluated, only some of the kale accessions, mainly leafy and short leafy kale group, exhibited high levels of resistance to clubroot, while all cabbage and cauliflower accessions were susceptible (Manzanares-Dauleux et al. 2000). Some of these resistant sources are widely used in the breeding program of *B. oleracea*. A detailed genetics of clubroot resistance was studied in *B. oleracea* by several groups and most of them concluded that the inheritance of this trait in *B. oleracea* is polygenic. Yoshikawa (1993), in a study involving progenies of a cross involving resistant cultivar Bohmerwaldkohl and a susceptible cabbage line, observed inheritance of resistance to be govern by one to four genes. Voorrips and Visser (1993), based on the genetic analysis of the three F1 populations derived from crosses between resistant lines Bohmerwaldkohl, Wisconsin, curly kale and susceptible cabbage line, reported recessive inheritance of
clubroot resistance. Laurens and Thomas (1993) observed that resistance to clubroot in kale is controlled by many alleles with additive effects. Based on qualitative and quantitative analysis of the $F_1$, $F_2$, and backcross progenies of four crosses derived from four different resistant sources and one common susceptible doubled haploid line, Voorrips and Kanne (1996) observed different mode of inheritance of resistance to clubroot. Of the four resistances studied, one was controlled by two complementary genes. Chiang and Crête (1976) agreed that the resistance was controlled by two loci by using the same resistant source.

Several insect pests damage *B. oleracea* crops, the most important of which are cabbage looper (*Trichoplusia ni*), cabbage worm (*Pieris rapae* L), cabbage aphid (*Brevicorne brassicae*), and diamond back moth (*Plutella xylostella*). Dickson and Eckenrode (1980) found that PI 234599 cauliflower having glossy leaves is immune to cabbage looper, cabbage worm and to diamond back moth. Genetic analysis showed that the resistance is recessive and quantitative and is linked to the glossy trait.

Tip burn, a major physiological disease in cabbage results due to non-translocation of calcium to the rapidly growing leaf tips, is characterized by browning and blackening of leaf tissues due to death of cells. Genetic analyses of resistance to tip burns by different study groups have contrasting views. Walker et al. (1965) reported that resistance was governed by two to three recessive genes, while Dickson (1977) gave evidence for dominance resistance controlled by two to three genes.

### 2.2.3 Early Linkage Analysis

Linkage analyses for several traits have been done in different studies involving crosses between different cultivar groups of *B. oleracea*. Linkage analyses or association of traits are important for designing breeding strategies in any crop. Linkages of genes for desirable traits are important while linkages of genes for favorable and unfavorable traits are not desirable for crop breeding. Economically important traits are often found to be associated with non-desirable traits in most of the agricultural crops, which hamper the breeding process. The most striking example is the non-occurrence of heading trait of cabbage with curled leaves of kale in $F_2$ or subsequent generations derived from crosses between heading cabbage having the entire leaf and non-heading kale with curly leaves. It is also observed that in progenies of such crosses head formation decreases when curliness increases and vice versa (Pease 1926; Allgayer 1928; Malinowski 1929; Samson 1930). When cabbage is crossed with kohlrabi and cauliflower, the heading trait of cabbage is not combined with the enlarged stem of kohlrabi and flower clusters of cauliflower in segregating progenies (Pease 1926). From this observation it is clear that genes from a particular botanical
variety or cultivar groups of *B. oleracea* tend to remain together and do not combine with the traits of other botanical variety. Malinowski (1929), Pease (1926) and Allgayer (1928) proposed linkage of different gene groups which are: genes for heading and sessile, entire and wide leaves; genes for heading and short stems and non-curved leaves; genes for heading, color (*P*), petiole blade, leaf wide, frilled leaves, and broad vs. narrow stem; and genes for heading, stem length, additional genes for leaf frilling, and the second factor for petiole blade. Observation of recombination between enlarged axil buds within cabbage wrapper leaves was also reported (Pease 1926; Allgayer 1928; Malinowski 1929; Detjen and McCue 1933). Independence of the gene for color has been reported from heading genes (Pease 1926). Recombination of high ascorbic acid content with that of resistance to cabbage yellows has been identified (Walker 1944; Foster and Walker 1946; Walker and Pound 1947).

However, certain percentages of trait recombination in progenies have been observed for some traits while crossing between different *B. oleracea* groups. Notable examples are observation of 20% crossing over between heading factor *K* with purple blade *P* gene by Allgayer (1928), 10% crossing over between the gene for petiolate *Pet* with *N* for non-heading, and 30% crossing over between *T* (tall plant) and *N* (*heading*) by Pease (1926). He also observed that the gene for purple leaf, *D*, is independent of recessive genes for heading, *n,n*. Kwan (1934) observed 10% crossing over between *R1* and *W* or *R2* and *Sm*. Other studies by Kristofferson (1927) and Malinowski (1929) showed a linkage between color genes *B* with either *A* or *C*, and genes for enlargement of stem in kohlrabi and color gene *D*, respectively. Recessive genes for bolting have been found to be independent of genes for heading and red color (Samson 1930; Detjen and McCue 1933).

Recently, Arus and Orton (1983) studied the inheritance and linkage relationships of variation in isozymes banding patterns for phosphoglucomutase (PGM), phosphoglucoisomerase (PGI), leucine aminopeptidase (LAP), alcohol dehydrogenase (ADH), acid phosphatase (APS) and glutamateoxaloacetate transaminase (GOT) in *B. oleracea*. However, no linkage relationship was observed between the six loci (*Adh-2, Aps-1L, Pgm-1, Pgm-2, Lap-1, and Aps-1*). Recent development in molecular markers particularly DNA markers have been successful in the development of molecular linkage map and linkage analysis of traits of economic importance and molecular markers, and identification, manipulation and cloning of many important traits in *Brassica* vegetables (Quiros 2001; Snowdon and Friedt 2004). The following chapters will be dealing with linkage analysis with molecular markers and traits QTL mapping and their manipulation in *Brassica* vegetable breeding.
2.2.4 Classical Breeding Achievements

Classical breeding and genetic analysis have been instrumental in the present day success of breeding *Brassica* vegetables for various traits. Today with the help of conventional breeding, desired quality and quantities has been achieved in *Brassica* vegetables. During the last few decades, extensive breeding efforts in vegetable brassicas had focused on improvement of agronomic traits, increasing yield, breeding for biotic and abiotic stresses and improvement in quality which are discussed below.

2.2.4.1 Yield and Agronomic Traits

Improvement of uniformity, shape, size, shelf-life and nutritional quality are part of vegetable *Brassica* breeding. In cabbage, breeding of the round shape with compact and solid frames has been achieved compared to earlier traditional varieties with a flat or pointed shape with a loose head and large frames. Because of this improved new type, a large number of plants can be accommodated in a unit area and can be harvested with machines (Dickson and Wallace 1986). Most of the present day cabbage cultivars are non-splitting and non-bolting compared to earlier premature bolting and splitting type. The most important quantitative traits for which success have been achieved in breeding are weight of the head, short growing period, early maturity, frost hardiness, long storage ability and morphological traits related to head development (de Moel and Everaats 1990). Head firmness and color, number of outer leaves, attractive green color of the wrapper leaves, and core dimensions have drawn major attention in breeding cabbage cultivars for commercial cultivation. Breeding for high harvest index by increasing the proportion of reproductive parts to the vegetative parts of the plants has been the major objective in cauliflower and broccoli breeding. Curd and head firmness has been achieved by reducing the floret stem lengths and grain size. This increases the post-harvest quality of the produce.

Another major achievement in vegetable *Brassica* breeding is development of uniform, heterotic and highly productive F$_1$ hybrids using self-incompatibility system. Of late, use of cytoplasmic male sterility for hybrid vegetable production has achieved major importance by transferring the male sterility system from *Raphanus sativus* into *B. oleracea*. This is more stable compared to self-incompatibility having multiple alleles that makes breeding more complex. Major seed companies these days are more interested in breeding F$_1$ hybrids because of the highly promising optimum market quality of hybrid vegetable brassicas. The curds/heads uniformity achieved by F$_1$ cauliflower and broccoli are in high demand for the freezing industry (Acciari et al. 1997).
2.2.4.2 Disease Resistance

Classical genetic analysis and screening of *B. oleracea* germplasms have identified many sources of disease resistance with varying degrees, some which have been exploited in resistance breeding programs. Several cauliflower cultivars resistant to *Xanthomonas campestris* have been selected in India and genetic analysis found that this resistance is dominant and polygenic in several Indian landraces (Bianco and Pimpini 1990). Notable examples of cauliflower resistant lines for this disease are Sn 445, Pua kea and MGS2-3 (Sharma et al. 1972). Pusa Shubhra was developed using Pua kea and MGS2-3 lines and recommended for commercial cultivation (Singh et al. 1993). Walker and Larsen (1980) identified two forms of pathotype-specific major resistance genes for clubroot diseases in cabbage while Weisaeth (1977) reported recessive resistance to this disease in the fodder turnip cultivar “Bohmerwaldkohl” determined by more than one gene. Resistance found in this cultivar has been transferred to many vegetable brassicas. Many of the important cabbage cultivars in the US are now resistant to cabbage yellows disease (Dickson and Wallace 1986). Early purple cauliflower was reported resistant to yellows. Sclerotinia rot caused by *Sclerotinia sclerotiorum* is a major disease of cauliflower in the early stage of development. Several cultivars with moderate resistance have been reported i.e. EC131592, Janavon, EC103576, Kn-81, early Winter Adam’s White head, EC162587, EC177283 (Kapoor 1986; Baswana et al. 1991; Sharma et al. 1995, 1997; Singh and Kalda 1995). Pusa Snowball K-25 has been developed from EC103576 as a resistant source and Pusa snowball-1 is released for commercial cultivation. Resistance to downy mildew disease has been identified in many cauliflower lines, e.g., Igloo, snowball Y, MGS2-3, 1-6-1-4 (Kontaxis et al. 1979; Chatterjee 1993). Pusa Hybrid 2 and Pusa Snowball K-25, resistant to downy mildew were released for commercial cultivation in India.

*Alternaria brassicae* or *Alternaria brassicola* also causes brown to black, small elongated spots in leaves, stems, and inflorescences of brassica plants. In cauliflower it causes curd blight and resistance to this disease was found in Indian cauliflower lines MGS2-3, Pua kea and 246-4 (Sharma et al. 1991), IIHR 142, IIHR 217 (Pandey et al. 1995) and snowball KT-9 (Sharma et al. 1991). Commercial cultivar Pusa Shubhra released in India is resistant to curd blight (Singh et al. 1993).

2.2.4.3 Insect Resistance

Several insects belonging to specialist and generalist groups attack vegetable brassicas and damage leaves, roots, stems and other parts of the plants making the crop unsuitable for human consumption. Breeding resistant
cultivars of *B. oleracea* to insect pests to reduce major economic loss was also a major objective among vegetable *Brassica* breeders. Several studies identified resistant sources for different insect pests (Radcliffe and Chapman 1966; Dickson and Eckenrode 1980; Lamb 1989). Resistance to *Pieris rapae*, *Plutella xylostella* and *Mamestra brassicae* has been identified in some cabbage lines of the USA (Dickson and Eckenrode 1975, 1980; Dickson and Wallace 1986). Radcliffe and Chapman (1966) have reported the differences in cultivar response to infestation of various insects in the field condition. Stoner (1990, 1992) reported that glossy leaves and high concentrations of glucosinolates are preferred leaf traits for oviposition of some insects. Red cabbage cultivars are less preferred for oviposition than green cultivars (Stoner 1992). Dickson and Eckenrode (1980) reported a glossy leaved cauliflower PI 234599 to be resistant to lepidopteron insects, diamondback moth, cabbage looper, and imported cabbage worm. Cabbage head borer resistance had been reported in many cultivars and F1 hybrid cauliflowers. The notable examples are Katiki, Early Patna, and F1 s aa × ES102, aa × Katiki (Lal et al. 1991, 1994). Resistance to cabbage aphids have been identified in cauliflower lines NY 13816, NYIR 9605 and others. Picoaga et al. (2003) found resistance to *M. brassicae* in Galician kale landraces.

### 2.2.4.4 Quality Traits

*Brassica* vegetables are known widely for the presence of rich nutrients and chemical components such as calcium, carotene, vitamin C, vitamin E and a high amount of antioxidants besides providing lots of vegetable fiber (Cao et al. 1996; Nilsson et al. 2006). During the last few decades extensive efforts have been made for breeding nutritionally high quality *B. oleracea* cultivars. Glucosinolates, sulfur-containing secondary metabolites is commonly found in the members of Brassicaceae family. Three types of glucosinolates have been named depending on the amino acid precursor, i.e. aliphatic derived from methionine, aromatic derived from tyrosine or phenylalanine and indolyl derived from tryptophan (Giamoustris and Mithen 1996). Of the three types, the most common in *Brassica* crops are aliphatic glucosinolates. Upon tissue damage, the breakdown product of these glucosinolates gives isothiocyanates, thiocyanates, epithionitriles, oxazolidone and nitriles (Fenwick et al. 1983; Halkier and Gershenzon 2006). Glucosinolates are known for their role in plant defense against fungi, nematodes, herbivores, and weeds (Rosa et al. 1997). Recently, isothiocyanates derived from glucosinolates breakdown products have been found to be anticancer, which induces Phase II detoxifying enzymes such as Quinine reductate, Glutathione-S-transferase, and others (Zhang et al. 1992; Brown et al. 2002). Sulforaphane, the isothiocyanate derived from glucoraphanine found in broccoli is an important subject of study for
the last decade. Breeders are trying to increase the glucoraphanine content in broccoli so that this functional food can prevent the risk of cancer in the human body. Genes involved in glucosinolate biosynthesis pathway have been cloned and characterized in *Arabidopsis thaliana* and in brassica crops (Li and Quiros 2002; Halkier and Gershenzon 2006; Bisht et al. 2009).

2.2.4.5 Transfer of Cytoplasmic Male Sterility into *B. oleracea*

Cytoplasmic male sterility is widely used in many crops species for using as a pollination control mechanism for hybrid seed production. Since, there is not any naturally occurring cytoplasmic male sterility in *B. oleracea*, cytoplasmic male sterility from *B. nigra*, *B. napus* and related Cruciferae species radish were transferred by many researchers (Pearson 1972; Bannerot et al. 1974; Dickson 1975; McCollum 1981; Chiang and Crete 1987; Hoser-Krauze 1987; Crisp and Tapsell 1993). Cytoplasmic male sterility “ogura” found in radish by Ogura (1968) was first transferred to broccoli by backcrossing (Bannerot et al. 1974; McCollum 1981). This was later transferred from broccoli to cauliflower by Dickson (1975) and Hoser-Krauze (1987). The transfer of this cytoplasmic male sterility was done in heat tolerant Indian cauliflower such as MS-91, MS-51 and is being used in F1 hybrid seed production (Sharma et al. 2005). In cabbage *B. nigra* cytoplasmic male sterility was transferred by Pearson (1972), while Chiang and Crete (1987) transferred *B. napus* cytoplasmic male sterility into it. Crisp and Tapsell (1993) later transferred *B. napus* cytoplasmic to make a sterility system from cabbage to cauliflower.

2.3 *Brassica rapa*

Many types of *B. rapa* vegetables were traditionally used as vegetables in East Asian countries including China, Japan and Korea. Chinese cabbage [*B. rapa* L. ssp. *pekinesis* (Lour) Olsson] is the most common type of vegetable in *B. rapa* eaten as fermented food “Kimchi” in Korea and as a fried vegetable in China. Pak choi [ssp. *chinensis* (L) Makino] is used as a fried vegetable in China and as a fresh vegetable of “Ssam” in Korea. Different from other *B. rapa* vegetables, turnip [ssp. *rapifera* (Metzy)] is used as a root vegetable. Root or the whole part of the turnip plant is consumed as “Kimchi” in Korea, as a fried vegetable in China and pickled in Japan.

2.3.1 History and Evolution of Different Types

The first record of brassica crop was mentioned in ancient Chinese literature, edited in the 5th century BC, as turnip (*B. rapa* ssp. *rapifera*) and mustard (*B. junceae*). From then to the 7th century, turnip was grown in north China
and pak-choi in South China. In the 10th century, at Yang-chou city located in central China, a vegetable was found having wrinkled, hairy leaves as large as fans and not so fibrous as pak-choi in texture. These characteristics are like the primary form of Chinese cabbage (Brassica rapa L. ssp. pekinensis). This may suggest that Chinese cabbage might have originated from natural hybridization between turnip and pak-choi. To prove this, artificial crosses were made between turnip and pak-choi collected from North and South China, respectively, from 1960 to 1962. The characteristics of the hybrids obtained were similar to that of Chinese cabbage, having large thin, hairy leaves with fairly clear wings along petioles. In the literature of the 12th century, the primary form introduced to the North developed thick petioles. After that many farmers tried to select more heading types, a premature form of Chinese cabbage was reported in a horticultural book in the 14th century. The typical form of Chinese cabbage was found in the 17th century at a local record of Shiu-Tian-Fu. Li (1981) classified Chinese cabbage into 11 types following morphological, ecological and economical characteristics using local cultivars in China.

In 2007, the cultivation area in China of Chinese cabbage was 26 billion m² followed by cabbage (8.9) and Pak-choi (5.3). The major cultivation provinces of Chinese cabbage were Shandong and Hebei, located in North China, covering 21%. However, in South east China, one of the most popular vegetables is pak-choi. Especially in Zhejiang province, where the cultivation area of Pak-choi was 0.9 billion m² while that of the Chinese cabbage was 0.3.

Pak-choi is a unique brassica vegetable in China. Thus traditional local varieties and high quality hybrid cultivars were cultivated together throughout the country. One popular inbred variety is “Aikang qing” possessing wide petioles and a high ratio of petiole weight and resistance to TuMV, which made this variety popular in most of China from the early 1980s (Fig. 2-1). Another popular inbred variety is “Shanghai qing”, which is very popular in Shanghai and later was cultivated throughout China (Fig. 2-1). This variety is taller than the “Aikang qing” and also shows a high quality of wide petioles and a high ratio of petiole weight. Although not as popular as the other two varieties, “Siyue man” (Wuyue man) possesses bolt resistance, and can be cultivated in spring (Fig. 2-1). Even though the taste of “Siyue man” is inferior, bolt resistance has made this variety popular in the northern provinces of China for winter and spring cultivation. High quality hybrid cultivars were also imported from other countries. One such popular cultivar in the Zhejiang province “Huaguan” was imported from Japan. This cultivar required a short period for harvesting and had a good quality. However its high price prevented further spreading of this cultivar. The official record of introducing Chinese cabbage in Japan was in 1875 at
Naito Shinjuku Testing Station. After the Russia-Japan war in 1905, Japanese soldiers brought seeds of Chinese cabbage and simultaneously several trading companies started to import various types of Chinese cabbage seeds. After 1920, Chinese cabbage seed production was established in Japan and this made it possible to develop various cultivars (Watanabe 1981).

In Korea, ancient literature of the 13th century mentioned the introduction of Chinese cabbage as a medicinal herb imported from China. In literature of the 17th century, cultivation of Chinese cabbage was mentioned as a vegetable using imported seeds. In the early 19th century, semi-heading type of the local variety was imported and produced a high quality Chinese cabbage in the north province of “Kae-sung” in Korea. This local variety became popular and got a name of “Kae-sung” Chinese cabbage. Using the seed propagation method established in 1850, the “Kae-sung” variety was cultivated in the central province of Korea. Especially in the province of Seoul, smaller but harder heading type was derived and got the name of “Seoul” Chinese cabbage (Fig. 2-2). In 1906, a government research station of Horticultural Demonstration Station was established that started research on vegetable breeding to overcome starvation. As Chinese cabbage and radish (Raphanus sativus L.) are the main vegetables for “Kimchi”, which is basic side dish in Korean cuisine, research focused on increasing yield. When Dr. Woo Jang-choon returned from Japan, hybrid breeding was started on Chinese cabbage and radish. Dr. Woo contributed to commercial breeding by developing valuable materials and educating students. In the 21st century, 880 varieties of Chinese cabbage, 1,040 varieties of radish, and 18 varieties of turnip were registered by the Korea Seed and Variety Service.
2.3.2 Breeding for Yield and Morphological Traits

Two major local varieties were planted in Korea before the 20th century. The characteristics of the first one, “Kae-sung” Chinese cabbage, are about 40 cm height, semi-heading type, about 10 leaf numbers with dark green colored curly leaf blade, and less hairy. The “Seoul” Chinese cabbage is shorter than the “Kae-sung” Chinese cabbage, semi-heading type, and has about 20 leaf numbers with a light green color. At first, the seeds of “Kae-sung” Chinese cabbage were imported from China, as the seed production failed in the Kae-sung province, where this variety was popular. After development of seed production using autumn harvested plants, this variety was favored and was planted as far down as the Seoul province. Farmers in Seoul continuously selected plants showing good characteristics for making “Kimchi” from the “Kae-sung” variety and developed the “Seoul” variety which was planted in many provinces located at the south of Seoul.

Breeding for yield increase was the most important purpose of the Horticultural Demonstration Station during the 20th century. The first step was introducing various exotic varieties and evaluation of their yield and quality. The planting period of heading Chinese cabbage was longer than semi-heading or non-heading type of local varieties. Farmers planted low quality semi- or non-heading type to produce twice a year before introducing early harvesting heading type in 1960. After this variety, introduction of various heading types (semi or non heading type), which had been traditionally planted in Korea over 800 years. Nineteen foreign
varieties were imported and seed production experiments were done on five selected varieties from 1953 to 1959. Late bolting high yielding heading type Chinese cabbage variety “Yaki” was developed from a Japanese cultivar for spring production of Chinese cabbage. Dr. Woo started making hybrid cultivars using self-incompatibility (SI) in Chinese cabbage. In 1960, he developed the first hybrid cultivar “Wonye 1ho” possessing characteristics of disease resistance, tight heading, and high yield (Fig. 2-3). The amount of seed production was 38 kg/ha using SI in a remote place to prevent pollen contamination. This hybrid cultivar development and seed production methods were given to several private seed companies that had the facilities and labor required for seed production. After this cultivar, numerous other varieties and inbreds were developed and distributed to private seed companies that boosted up seed industries in Korea.

Figure 2-3 Hand drawings of “Wonye 1ho” (left) and “Wonye 2ho” with parental materials (right).

2.3.3 Heterosis Breeding

The main purpose of the hybrid cultivar was high yield and year round production of Chinese cabbage after 1960. The first hybrid cultivar was “Wonye 1ho” for autumn cultivation followed by “Wonye 2ho”, distributed from the Horticultural Demonstration Station to seed companies and farmers (Fig. 2-3). The main characteristic of these cultivars was big head size thus requiring a long planting period. The sowing time was before August 5 at the central province at that time, which witnesses viral disease and soft rot due to high temperature. In late autumn, the temperature drops sharply thus sowing time could not be delayed using these cultivars.

The best way of obtaining good hybrid cultivar is making crosses between inbreds and evaluating horticultural traits. However, this specific combining ability test needs a lot of labor and money. If a breeder have 100 inbreds, 5,050 times of bud pollinations are required only for making seeds.

In 1971, a breeder developed one method of making a combination between two parental groups. The horticultural traits of each group were classified by inspecting two parents of “Wonye 2ho” distributed from the
Classical Breeding and Genetic Analysis of Vegetable Brassicas

National Horticultural Research Institute (NHRI) in 1960. The A group has characteristics of few leaf numbers, big outer leaves, and incomplete heading, thus can be easily discarded as inferior inbreds. The B group has characteristics of short height, profuse leaf numbers, and similar leaf size, fast growth and complete hard heading. As the head is small and tends to grow ugly, these inbreds were also discarded. However, when hybrid seeds were produced involving these two groups, the horticultural traits were excellent as compared to other crosses using both nice looking inbreds as parents. The first commercial hybrid developed by this crossing method was “Dea-hyung-Garak”, an autumn cultivar launched in 1971 and has been planted until recently.

One parental material of “Dea-hyung-Garak” was derived from the early harvesting “Cheong-Bang” that possesses viral disease tolerance and was distributed from a seed company. The other parental material was derived from a Japanese cultivar “Hirathuka 1go” possessing characteristics of early harvesting and strong SI. In the early 1970s, cultivars could be harvested at least 80 days after sowing when the big and half-folding head matured. “Dea-hyung-Garak” could be harvested 70 days after sowing when the medium sized rose shape heading was formed. This beautiful head shape attracted customers, reduced viral disease infection by late sowing and resulted in a short planting period that attracted the farmers. Thus, this cultivar replaced other big semi-folding headed cultivars and became popular.

One seed company continued developing the rose-shape heading variety while other seed companies focused on semi-folded heading type (Fig. 2-4). As a result of continuous breeding in the commercial seed companies and the government research stations, farmers could now select what they wanted from among various high quality hybrids of Chinese cabbage.

Figure 2-4 A rose shaped head (A) and semi-folded head (B) of hybrid cultivars in Chinese cabbage.

Color image of this figure appears in the color plate section at the end of the book.
DNA markers such as randomly amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), simple sequence repeats (SSR), and sequence tagged site-polymerase chain reaction (STS-PCR) facilitate the major technique for genetic distance estimation in crops. Among them, RAPD markers are commonly used in a large number of species (Williams et al. 1990). This may be due to the fact that the RAPD facilitates automation, and the low cost of operation; besides, the procedure is simple, speedy, and shows optimum degree of polymorphism in addition to wide genome coverage (Dias et al. 2003).

As the RAPD markers are non-locus specific and low reproducible (Kesseli et al. 1994), several studies failed to use these markers in determining relationships between markers and traits or heterosis (Joyce et al. 1999; Liu et al. 1999; Kwon et al. 2002). However, many studies reported the successful results with RAPD markers. The RAPD distance calculated between species explains the general combining ability and the specific combining ability of the phenotypic character with a global coefficient of determination of 81.6% between two species of *Eucalyptus*. From 26 samples, 415 reliable RAPD bands were obtained using 15 primers. Among them, the effective RAPD markers for predicting hybrid performance were those presenting common frequencies in the two species of *Eucalyptus* (Baril et al. 1997).

Even though many papers have recently reported using DNA markers to predict the combining ability, commercial breeders prefer using the traditional grouping method developed in 1971 to make a combination.

In the 1960s, breeders used the bud pollination method in SI plants of vegetable Brassica crops for development of inbred lines. They counted the number of seed set per flower after self pollination to find out the strength of SI. When an inbred line set seeds less than five per flower, it was considered to possess strong SI and was selected as parental material. The selected parental Inbreds were propagated using artificial bud pollination in a net house to prevent contaminating bees or insects.

The most difficult problem in breeding “Dea-hyung-Garak” was selecting inbreds showing strong SI derived from “Hirathuka 1go”. After failure of the first screening using tens of inbreds, the breeders used two big size net houses to screen hundreds of inbreds and finally selected one inbred. Because the “Hirathuka 1go” was a hybrid between Chinese cabbage and cabbage to introduce soft rot disease resistance in Chinese cabbage, the breeders figured that the “Hirathuka 1go” was not yet genetically fixed.

Self-incompatibility (SI) prevents self-fertilization and promotes out-crossing in hermaphrodite plants (Nettancourt 1977). In brassica crops, SI is sporophytically controlled by a single, highly polymorphic S-locus consisting of a series of multiple alleles (Bateman 1955). Physical mapping and sequencing analysis of S-haplotypes in *Brassica* species have identified three highly polymorphic major SI genes at the S-locus. The *SRK* and *SPII/SCR*
genes determine the $S$-specificity of the stigma and the pollen, respectively (Schopfer et al. 1999; Takasaki et al. 2000; Takayama et al. 2000; Shiba et al. 2002). The $SLG$ gene encodes an abundant, soluble, secreted glycoprotein that enhances the SI recognition reaction perhaps by stabilizing the SRK (Roberts et al. 1979; Dixit et al. 2000; Takasaki et al. 2000). The SRK gene consisted of a highly polymorphic $SLG$ domain, which is very similar to the $SLG$ sequence, a trans-membrain domain and conserved protein kinase domain.

A gene encoding a small cysteine-rich protein, termed $S$ locus cysteine-rich protein (SCR) was identified between $SLG$ and $SRK$ genes in the course of genome analysis of the $S_8$ haplotype of $B. campestris$ (Schopfer et al. 1999). The analysis of the $S_8$ haplotype around the $SLG/ SRK$ region by two independent groups clearly demonstrated that SCR is allelic to $SP11$. Thus, the gene encoding this small cysteine-rich protein was referred to as $SP11/ SCR$ (Schopfer et al. 1999; Takayama et al. 2000).

Both $S_9$ and $S_8$-$SP11/ SCR$ contain eight conserved cysteine residues and relatively conserved putative signal peptides, but most regions of the mature protein exhibited low homology. Using a primer designed from the conserved signal peptide sequence, it was possible to amplify allelic $SP11/ SCR$ genes from most of the pollen-dominant $S$-haplotypes class-I, (Watanabe et al. 2000) and the pollen-recessive $S$-haplotypes class-II (Shiba et al. 2002).

The variation of SI activity due to both genetic and environmental factors was observed. Some cabbage lines showed different levels of SI activity between genotypes/lines, and from plant to plant even within the same lines (Wallace 1979). The method of flower bud pollination was used at first to examine the SI genotype. The result of flower pollination was reliable but it needed a long time and involved the risk of contamination.

The SI activity was reported to change by temperature and humidity. Thus, a method of investigating the activity of SI without interference of environment was developed in brassica vegetables (Wallace 1979; Lee et al. 1982). In vitro pollination with 98% relative humidity proved to be the most reliable as compared to greenhouse pollination. Counting the number of penetrated pollen tubes using decolorized aniline blue as a fluorescent stain was a simple and reliable method of assigning the activity of SI allelic phenotype.

There are some reports that the SI activity was not only controlled by the $S$-haplotype but also by other genes. A single nucleotide deletion or alternative transcripts of the $SKR$ genes located at the $S$-locus resulted in a low SI activity (Göring et al. 1993; Tantikanjana et al. 1993). The activity of SI of $S$-haplotypes was identified with seed set analysis in 88 lines of radish. The activity of SI appeared to differ depending on the genetic background of inbred lines. It was expected that existence of other factors related to the SI activity is possible (Seo et al. 2003).

The $M$ (modifier) gene was located at a locus that is not linked, and recessive mutation ($m$) is epistatic to the $S$ locus. It functions in the regulation of water availability at the papilla cell surface, which would be required
for the SI reaction as a component of the SRK signaling pathway (Ikeda et al. 1997). A recessive mutation of the M gene eliminates the SI response in the stigma. Positional cloning of M revealed that it encoded a membrane-anchored cytoplasmic serine/threonine protein kinase, designated M locus protein kinase (MLPK). Transient expression of MLPK restores the SI response of mm papilla cells, suggesting that MLPK is a positive mediator of brassica SI signaling. The plasma membrane localization of MLPK suggests that it might stimulate the function of SRK (Murase et al. 2004).

The seeds of inbred lines having strong SI were propagated with over 3% CO₂ treatment at night following bee pollination during the daytime. The CO₂ treatment required a specific plastic house, which can seize high concentration of CO₂, a CO₂ supply system, bees as pollinators, and ventilation labor for the active pollination of bees (Lee 1979). With this method, the strong SI activity of parental lines was nullified for a while, which made it possible to produce seeds of parental lines in Chinese cabbage. With the help of high concentration CO₂ treatment, the single crossed cultivars, showing high uniformity, have been commonly used for breeding hybrid seeds of brassica vegetables in Korea.

2.3.4 Breeding for Quality

The first cultivar possessing yellow inner leaf was “Norang spring” launched in 1988. This cultivar showed semi-folding heading type and disappeared about 20 years ago because the more favored variety “Dea-hyung-Garak” with a rose shaped heading type was developed. This variety was most popular for around two decades among growers and in the markets. However, when “Norag spring” was launched, the yellow inner leaf color distinguished this from others and customers wanted to buy it. The taste of this cultivar was good and customers believe that the yellow inner leaf of Chinese cabbage tastes like roasted sesame (Fig. 2-5). Furthermore, this cultivar is bolting resistant and not sensitive to calcium deficiency compared to the pre-existing cultivar “High land summer”. After “Norang spring” became popular, the characteristic of the yellow inner leaf was introduced not only into other spring cultivars but also into autumn and winter cultivars. Now in the 21st century, even high school students in Korea think that the natural color of Chinese cabbage is yellow.

2.3.5 Breeding for Abiotic Stress Tolerance

Chinese cabbage has tolerance to low temperature but is sensitive to high temperature. Thus for autumn cultivation sowing in late summer and harvesting in late autumn are most natural and can produce high quality Chinese cabbage. However the demand for “Kimchi”, especially made of
Chinese cabbage, occurs all year round, forcing breeders to develop heat tolerant cultivars, which can grow enough head in hot weather. In 1973, a heat tolerant summer cultivar “Nae-Seo-beak-ro” was developed by a commercial seed company and this made it possible to grow Chinese cabbage in summer at high lands. One parent of this cultivar was collected from the National Horticultural Research Institute (NHRI), renamed as the Horticultural Demonstration Station, and the other parent was an inbred derived from the Japanese spring cultivar “Mu-so” developed by a seed company. “Nae-Seo-beak-ro” could grow well and make heads at high temperature, could endure high humidity in the monsoon, and it showed resistance to viral disease, soft rot and downy mildew. Using this cultivar in highlands, farmers could harvest considerably high quality Chinese cabbage in summer. After this cultivar, many summer cultivars possessing high heat and disease tolerance were developed by the seed companies.

In the four seasons, the temperature in spring is considerably conducive for growing Chinese cabbage. However, the low temperature of early spring causes bolting that reduces the quality so that Chinese cabbage cannot be used for “Kimchi”. Thus breeders tried to develop a bolting tolerant cultivar. In the early 1970s, seed import of a Japanese cultivar “Chun-Chu”, popular as a high land summer cultivar in Korea, was impossible and this caused a social problem. At that time, a heat tolerant late bolting cultivar “High Land Summer” was developed by a seed company in 1976 to solve this problem. One parent of this cultivar was an inbred derived from the Japanese cultivar “Kyoto 2go” collected from NHRI and the other parental inbred was derived from the Japanese cultivar “Chun-Chu”. As the name represents, “High Land Summer” was developed for summer cultivation in highlands. In highland summer cultivation, unexpected low temperature caused bolting in Chinese cabbage. To select late bolting inbreds, breeders lowered the temperature to 5–10°C during 20 days when the plant reached six true leaf stage and then grew at a natural high temperature of summer. Using this method, they could select enough number of high quality inbred lines to make various crossings and successfully developed a hybrid cultivar.
possessing a rose heading shape, heat and disease tolerance, and good taste. As this cultivar possessed a late bolting characteristic, farmers noticed that this cultivar can be planted in spring. Later this “High Land Summer” cultivar was widely used as a spring culture nationwide. The “High Land Summer” cultivar is the second long-lived popular cultivar in Korea lasting over 20 years and planted until recently.

Winter culture was possible only in Jeju Island during the 1980s using the Japanese cultivar “Winter”. In the late 1980s, the south-coast province imported the new Japanese cultivar “Snow Wind” for winter production of Chinese cabbage. In 1988, a Korean seed company launched a winter cultivar “Ha-ru-bang” for Jeju Island. However, this cultivar could not endure the cold winter conditions and disappeared soon. In 1992, a new winter cultivar “Dong-pung” meaning “east wind” was developed by another seed company. This cultivar grew well in cold winter with a strong root, possessing strong resistance to calcium deficiency and bolting, and tasted good. As this cultivar was highly resistant to cold temperature, cultivation in the south-coast area could produce high quality Chinese cabbage in winter using barren land and the produce could easily be transferred to many big cities in Korea (Fig. 2-6). Thus, winter cultivation in Jeju Island requiring high transportation cost could be minimized.

With the help of developing heat tolerant, late bolting and cold tolerant hybrid cultivars, respectively, Korean farmers could harvest high quality Chinese cabbage while conveniently using barren land all year round. Furthermore, Koreans can enjoy tasty fresh Chinese cabbage at a low price in hot or cold seasons.

Figure 2-6 Winter cultivation area in South-coast province of Korea.
Color image of this figure appears in the color plate section at the end of the book.
2.3.6 Breeding for Biotic Stress: Diseases

In the 1970s, the most severe disease that caused reduction was viral disease in Chinese cabbage. One commercial breeder selected inbreds and F1 hybrids using special fields where severe viral diseases occurred every year. He launched a cultivar of “Peyongang sin 2ho”, which could endure severe viral disease occurring in the southern province of Korea. The Asian Vegetable Research and Development Center (AVRDC), now renamed as The World Vegetable Center, collected worldwide Chinese cabbage germplasm and inoculated them with turnip mosaic virus (TuMV) during 1977–1979. Using five strains of TuMV for artificial inoculation, only “Peyongang sin 2ho” was found resistant to every strain.

When the nursery cultivation method was established, insect-mediated viral disease could be controlled in the early growing stage of Chinese cabbage. These days farmers who want to sell the product, transplant young seedlings in their production field to harvest high quality Chinese cabbage safely. Thus, the problem of viral disease has disappeared after using these seedlings.

In the 20th century, clubroot disease occurred in all areas at every cropping season. The first occurrence of clubroot was reported at Gyeonggi province in 1992 (Lee et al. 2001). From 2002, in every area of the Korean peninsula, including Jeju Island, clubroot disease was found to occur in Chinese cabbage.

The first cultivar showing clubroot resistance was “CR singsing” launched in 1995. Farmers in the Kangwon province, famous for alpine summer culture, faced a severe problem of clubroot mass occurrence because the cropping condition is suitable for the propagation of the clubroot germ. The clubroot resistance of “CR singsing” was introduced from the Japanese cultivar, “CR Shinkki”. At first, “CR singsing” was resistant in alpine summer culture; however after two years of cultivation, this variety became susceptible to clubroot.

Later in 2003, another variety, “CR green”, was developed by another seed company. This cultivar proved to possess not only clubroot resistance but also the powdery mildew resistance gene. However, it could not be widely cultivated because of the tough leaf texture, which reduced the efficiency of “Kimchi” producing factories. However, clubroot resistance of this cultivar was broken after two years of cultivation.

Recently, breeders are trying to introduce various clubroot resistance genes into many cultivars for spring, summer, autumn and winter cultivation. The resistance sources were introduced from Japan and distributed from the National Institute of Horticultural & Herbal Science (NIHHS), renamed NHRI. In NHRI, researchers are trying to find germplasm showing strong resistance to clubroot race 4 after a single spore isolation. Over 200 clubroot-
inoculated Chinese cabbage roots were collected from Korea and each race was identified using the WCD method. Among 13 identified races, race 4 was found most frequently and showed severe symptoms (Cho et al. 2001). Thus single spore isolate of race 4 was used to select resistance of *B. rapa* germplasm. Using nutrient solution for cultivating inoculated plants (Park et al. 2008), an accession of IT033820 was selected as resistance turnip and distributed to seed companies.

Until now, resistant hybrid cultivars are turning sensitive in some parts of Korea and this forced breeders and researchers to develop highly resistant but tasty cultivars in Chinese cabbage. This phenomenon occurs and causes severe problems worldwidely, thus joint research between nations would provide short-cuts to researchers and breeders (Fig. 2-7 and 2-8).

**Figure 2-7** Procedure of inoculation method of clubroot in Chinese cabbage (A; Seed germination, B; Input 20 mL of inoculum, C; Trans planting 15 day-old seedlings using 50 pot plug tray and urethane sponge.

*Color image of this figure appears in the color plate section at the end of the book.*

**Figure 2-8** Hydroponic culture of clubroot inoculated Chinese cabbage and the symptoms. (A; 2 weeks after transplanting, B; 6 weeks after transplanting, C; Symptoms of clubroot, left; sensitive, middle; medium sensitive, right; resistance).

*Color image of this figure appears in the color plate section at the end of the book.*

### 2.4 Limitations of Classical Breeding and Need for Molecular Breeding

Although classical vegetable *Brassica* breeding has been successful in many aspects such as breeding for plant morphotypes including shape, size, firmness, uniformity, colors and others, improvement in majority of the areas could not be achieved due to many factors. Genetic analysis to
Classical Breeding and Genetic Analysis of Vegetable Brassicas

determine the number of genes and gene action could largely be achieved for easily scorable qualitative traits, while the majority of economically important traits such as head and curd size of cabbage and cauliflower are quantitative in nature. Breeding for such kind of traits was largely complex due to polygenic inheritance and the influence of the environment. Further phenotypic evaluation is the hardest part of any breeding program, which not only demands extensive labor but also lots of time. F1 hybrid breeding in cabbage and cauliflower has relied on the self-incompatibility gene having multiple alleles with a complex inheritance pattern. This system in most of the cases was unstable and many of the F1 hybrids were of low quality due to breakdown of self-incompatibility. However, extensive efforts have been made to transfer available cytoplasmic male sterility system from *B. nigra*, *B. napus* and *Raphanus* into *B. oleracea* for hybrid breeding.

Many of the resistance sources identified in vegetable brassicas for diseases and insect-pests could not be transferred into suitable cultivars because most of the resistances identified were not completely resistance (Ordas and Cartea 2008). Notable examples are a complex inheritance pattern of resistance to black spot and clubroot disease identified in cabbage, cauliflower and kales. However, quantitative inheritance complicates the transfer of resistance genes in commercial cultivars. The case is similar in clubroot resistance breeding in *Brassica* vegetables. Many race specific resistances have been identified for clubroot in different sources, but exploitation of these genes in breeding could not be done due to incomplete resistance.

Breeding for quality of product by increasing the nutrient contents (vitamin C, carotenoids, antioxidants, storability) and functional foods (glucosinolates) have been major objectives in recent years along with increasing yield, resistance to diseases and pests and other agronomic traits. However, many nutritional compounds such as carotenoids and secondary metabolites such as glucosinolates are synthesized in complex biosynthetic pathways involving many genes. Conventional breeding for desired level of particular compounds is very complex and difficult.

Recent advances in molecular biology, particularly the development of molecular markers and transformation technology, have been very promising for easy identification and manipulation of qualitative and quantitative traits in economically important crop plants including *Brassica* vegetables (Snowdon and Friedt 2004). Molecular markers can locate and determine the number of genes controlling the traits of interest in the genome and using the linked markers, manipulation of traits can be done without much difficulty. This not only makes it easier but also saves the time required for selection of a particular trait without waiting for the actual time of expression of the traits. At the early stage, using linked markers one can select the genome and manipulate in breeding. Several molecular markers have been developed in *Brassica* species and molecular linkage maps have
been constructed. Highly saturated maps with co-dominant markers such as simple sequence repeat (SSR) would help in identification of some important recessive genes governing resistances to pests and diseases and quality traits in heterozygous conditions and hence enhance the selection process at the early stage without waiting for the time of character expression.

Agronomically important traits that were difficult to breed conventionally, have been mapped in the genome. Various genes and quantitative trait loci (QTL) controlling resistance against diseases and insect-pests have been mapped in vegetable Brassica genome using molecular markers (Snowdon and Friedt 2004). Molecular markers and genetic maps have been widely used to assist breeders in marker-aided selection in Brassica species. The genetic maps developed by using molecular markers are useful for the study of genome structure, evolution, mapping and tagging of important genes. In B. oleracea and B. rapa, linkage maps have been extensively used to identify economically important complex polygenic traits and efforts are being made to transfer these genes using the linked markers to desirable commercial background. Notable examples are mapping of disease resistance and male cytoplasm male sterility restorer genes (Snowdon and Friedt 2004). Linked markers to these economically important traits could be used to transfer this trait into commercial cultivars through marker-assisted selection, molecular manipulation of QTL/gene(s), and cloning and characterization of important genes. It is due to the construction of molecular genetic maps, that we can learn about the complex structure of Brassica genome including duplication, and intra- and intergenomic conservation of different chromosomal blocks. Brassicas are the close relatives of model plant Arabidopsis thaliana, whose complete genome sequence is available. Most of the genes have been characterized in Arabidopsis and molecular mapping and comparative alignment of Brassica genome with that of A. thaliana chromosomes showed conservation of chromosomal blocks. Further, at the gene level, conservation of 75–80% sequence identified has been reported between coding regions of different brassicas and A. thaliana. Comparative alignment of the Brassica chromosome containing genes/QTLs of interest would be helpful for identification of candidate genes from A. thaliana. This would help faster cloning and characterization of candidate genes of interest and manipulation in brassica breeding. Taking advantage of this mapping and tagging of glucosinolates genes has been done in Brassica juncea (Bisht et al. 2009). Gene-specific markers for high and low allele have been developed and are being used in B. juncea breeding for development of canola quality B. juncea.

Transgenic technology has been used to develop male sterility in B. napus and B. juncea using barnase and barstar system (Mariani et al. 1990; Jagannath et al. 2001; Bisht et al. 2007). These systems can be used for developing male sterile line in brassica vegetables for hybrid seed production because most of the cytoplasmic male sterility transferred
from *B. nigra*, *B. napus* and radish were associated with morphological abnormalities. Many *Bacillus thurienensis* genes, e.g., *Cry 1A*, *cry 1Ac*, *cry 2A*, have been transformed in *Brassica* species for developing insect-pest resistance (Sharma et al. 2005; Ordas and Cartea 2008). Likewise, many genes such as chitinase, glucanase, permantis, etc. can be transformed into *B. oleracea* for producing fungal resistant brassica crops. Many of the insect or disease resistance traits are not found in *Brassica* vegetables. In such cases the trait of economic importance should be transferred from related species through wide hybridization and the embryo-rescue technique. This technique has been used by many to transfer cytoplasmic male sterility from *B. napus* and radish (Inomata 1977; Ayotte et al. 1987). Microspore culture is also a very important technique for rapidly producing homozygous lines with diverse phenotypic variation from wide-crosses because vegetable brassicas are highly cross-pollinated because of self-incompatibility. Doubled haploid lines produced though microspore culture in *Brassica* vegetables have been used for development of linkage maps, mapping and tagging of many important genes and traits QTLs, and other genetic studies. Thus, breeding of brassica vegetables for desired plant type and quality will be greatly enhanced by using the above molecular tools and techniques. This will not only speed the breeding process but also help in precision breeding of vegetable brassicas.

**Aknowledgements**

This work was supported by grants for the Technology Development Program for Agriculture and Forestry (Grant no. 607002-05), Ministry of Agriculture, Forestry and Fisheries, Republic of Korea.

**References**


Classical Breeding and Genetic Analysis of Vegetable Brassicas


Kristofferson KB (1921) Untersuchungen iber die F1und F2, generation yon einem spontanen Bastard zwischen Weisskraut and Griinkohl. Sveriges Utsad Tidsk 31–52.


Kristofferson KB (1927) Contributions to the genetics of Brassica oleracea. II. Hereditas 9: 343–348.


Classical Breeding and Genetic Analysis of Vegetable Brassicas


Stoner KA (1992) Density of imported cabbageworms (Lepidoptera: Pieridae), cabbage aphids (Homoptera: Aphididae), and flea beetles (Coleoptera: Chrysomelidae) on glossy and trichome bearing lines of Brassica oleracea. J Econ Entomol 85: 1023–1030.


Diversity Analysis and Molecular Taxonomy of Brassica Vegetable Crops

Guusje Bonnema,1,3* Dunia Pino Del Carpio,1,4 and Jianjun Zhao1,2

ABSTRACT

Brassica crops have great economic value, harbor an enormous diversity and have adapted to cultivation in almost every part of the world. The triangle of U depicts the relationships between the three diploid species B. rapa, B. oleracea and B. nigra and the three amphidiploid species (B. napus, B. juncea and B. carinata) that resulted from interspecific hybridizations. In this chapter, first the different Brassica crops and radish are introduced, and their taxonomic position is described. Thereafter the literature on morphological and genetic diversity of vegetable Brassica’s that belong to the diploid species B. rapa (A genome), B. oleracea (C genome) and the amphidiploid species B. juncea (A and B genomes), and radishes that belong to the related species Raphanus sativus, is reviewed. Since these species are not only cultivated for vegetable crops, but also for fodder, oil and condiments, these crop types are also included in this chapter. For the four different species that are reviewed in this chapter, overviews of the origin, the crop types, and the diversity of cultivated and wild forms are given. In the final paragraph the use of genetic and phenotypic diversity assessments for germplasm conservation and plant breeding are discussed.

Keywords: Brassica rapa, Brasica oleracea, Brassica napus, Brassica juncea, Raphanus sativus, genetic diversity, morpho-type, centers of diversity

1Laboratory of Plant Breeding, Wageningen University, 6700AA, Wageningen, the Netherlands.
2Horticultural College, Hebei Agricultural University, 071001, Baoding, China.
3Centre for Biosystems Genomics, 6700 AB Wageningen, the Netherlands.
4Graduate School Experimental Plant Sciences, 6708PB Wageningen, the Netherlands.
*Corresponding author: Guusje.Bonnema@wur.nl
3.1 Introduction

Brassicas have great economic value and encompass a great diversity of plants. They are grown as vegetables, fodder, sources of oils and condiments. The morphological variation present within Brassica species is enormous and from several species practically every part of the plant is used by man. This includes the leaves in crops like heading cabbages and the leafy types that do not form heads (pak choi, komatsuna, etc.), the terminal and axillary buds in cauliflower, broccoli, broccolietto, the seedpods in seed stalk mustard, the swollen stems in tait sai and kohlrabi, the swollen roots in turnips and swede and the seeds in all oil crops. In this chapter, the focus will be on three important Brassica vegetable crops that belong to Brassica rapa, Brassica oleracea and Brassica juncea, and on Raphanus sativus, the radishes. Since B. rapa, B. juncea and R. sativus also include morphotypes that are cultivated as oil crop and/or condiments, the diversity analyses also extends to the oil crops. For both B. rapa and B. juncea the diversity within and between morphotypes and the taxonomic relationships between the vegetable and oil types are discussed. Besides the genetic diversity in R. sativus, the phylogenetic relationships between radish and B. rapa/B. oleracea and B. nigra are also discussed. Table 3-1 gives an overview of the different subspecies within the three Brassica species and R. sativus classified according to the organs that are used for consumption and Fig. 3-1 depicts the different crop types within B. rapa. In the different paragraphs for each species the origin of the species is discussed, the crop types are introduced and an overview of studies on the genetic diversity within that species is presented. In the subchapters on B. oleracea, B. rapa and R. sativus diversity of wild related species that have played a role in the evolution of the different crop types is also discussed. Before we describe the different Brassica and Raphanus species, an overview is given of the taxonomic position of these genera/species in the Brassicaceae family.

3.2 Taxonomy of the Brassicaceae

The Brassicaceae is a large plant family (338 genera and 3,700 species) of major scientific and economic importance. The taxonomy of this group is difficult, which is partly due to convergent evolution in nearly every morphological feature used to define tribes and genera. Koch and Kiefer (2006) state that rather than simply considering Brassica and Arabidopsis thaliana as dicotyledonous model organisms, the whole Brassicaceae family can be used as an example to understand phylogeographic patterns (evolution both in space and time) in a broader context. Even though there is a continuous increase in the number of phylogenetic studies available for various cruciferous taxa of different taxonomic levels
**Table 3-1** Overview of different subspecies within three *Brassica* species and *Raphanus sativus* classified according to the organs that are used for consumption.

<table>
<thead>
<tr>
<th>Organs used</th>
<th><em>B. rapa</em></th>
<th><em>B. oleracea</em></th>
<th><em>B. juncea</em></th>
<th><em>R. sativus</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Leafy type—Heading</td>
<td>Chinese cabbage (<em>pekinensis</em>)</td>
<td>Chinese kale (<em>alboflagbra</em>)</td>
<td>var. capitata and var. involuta</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Kales (<em>acephala</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>White/red headed cabbage (<em>capitata</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Savoy headed cabbage (<em>sabauda</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tronchuda cabbage (<em>costata</em>)</td>
<td></td>
</tr>
<tr>
<td>Leafy types—Loose leaves</td>
<td>Pak choi (<em>chinensis</em>)</td>
<td></td>
<td>Large, green or purple leaves (<em>rugosa</em> and <em>latipra</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wutacai (<em>narinosa</em>)</td>
<td></td>
<td>Dissected leaves (var. <em>crispifolia</em>, var. <em>foliosa</em>, var. <em>leucanthus</em>, var. <em>linearifolia</em> and var. <em>multisecta</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Komatsuna (<em>pervoidis</em>)</td>
<td></td>
<td>strumiferous mustards (var. <em>strumata</em>)</td>
<td></td>
</tr>
<tr>
<td>Leafy type—Tilling</td>
<td>Mizuna &amp;Mibuna (<em>nipponinica</em> &amp; <em>japonica</em>)</td>
<td></td>
<td>var. <em>gemmifera</em> and var. <em>multiceps</em></td>
<td></td>
</tr>
<tr>
<td>Axillary buds</td>
<td></td>
<td>Brussels sprouts (<em>gemmifera</em>)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swollen stem</td>
<td>Caixin or Caitai (<em>parachinensis</em>)</td>
<td>Kohlrabi (<em>gongyloides</em>)</td>
<td>Swollen stem mustards (var. <em>Tumida</em>, var. <em>Tsatsai</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zicaitai (<em>purpuraria</em>)</td>
<td></td>
<td>Swollen shoot mustards (var. <em>crassicaulis</em>)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Broccolo, Broccoli raab, Cima di rapa (<em>rauve</em>)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 3-1 contd...*
<table>
<thead>
<tr>
<th>Organs used</th>
<th>Brassica species</th>
<th>Swollen root</th>
<th>Inflorescence</th>
<th>Oilseed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Swollen root</strong></td>
<td>B. rapa</td>
<td>Turnips</td>
<td>Cauliflower</td>
<td>Oilseed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(swollen</td>
<td>(kohlrabi)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>hypocotyl/root)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Root mustard</td>
<td>Seed stalk mustard</td>
<td>Oilseed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(var. napiformis)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Black radish</td>
<td>Other</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(var. nigra)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inflorescence</strong></td>
<td>B. rapa</td>
<td>Broccoli</td>
<td>Broccoli</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(var. italica)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cima di rapa</td>
<td>(var. ruvo)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Oilseed</strong></td>
<td>B. rapa</td>
<td>Turnips</td>
<td>Cauliflower</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(var. napiformis)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Black radish</td>
<td>Seed stalk mustard</td>
<td>Oilseed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(var. nigra)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3-1 contd...
Diversity Analysis and Molecular Taxonomy of Brassica Vegetable Crops

(see review Al-Shehbaz et al. 2006 and references therein), few studies give a comprehensive understanding of the whole family. Three family-wide molecular phylogenetic analyses, which resulted in a proposed new classification scheme to organize genera into tribes, greatly improved the understanding of the evolutionary relationships in the Brassicaceae, compared to early attempts that were mainly based on traits like fruit morphological characteristics (Al-Shebaz et al. 2006; Bailey et al. 2006; Beilstein et al. 2006). Based on these studies, most of the 338 genera were placed into 25 tribes that provide a reference for future phylogenetic studies (discussed in Schranz et al. 2007). The phylogenies that are based on a comprehensive phylogenetic analysis of 746 nrDNA internal transcribed spacer (ITS) sequences, representing 24 of the 25 currently recognized tribes, 146 genera, and 461 species of Brassicaceae, confirmed the monophyly of most of these new tribes (Bailey et al. 2006). All the above studies support the lineages I and II as first defined by Beilstein et al. (2006). Lineage I

Figure 3-1 Drawings of Brassica rapa plants that represent the different morphotypes. A. ?pak choi, B. Chinese cabbage, C. winter oilseed, D. turnip, E. wutacai, F. mizuna, G. yellow sarson, H. broccoletto. Drawings were made by Dunia Pino Del Carpio from pictures of plants of the B. rapa core collection grown in Wageningen.
contains Camelineae (to which Arabidopsis belongs) and lineage II contains the agronomically important Brassiceae tribe. The tribe Brassiceae is a monophyletic group and comprises around 240 species and 49–54 genera. This tribe includes the economically important Brassica crops and radish (Raphanus). In the paragraph describing genetic diversity of R. sativus, the relationship between radish and the Brassica species will be discussed.

3.3 Diversity of Brassica rapa Vegetables and Oilseeds

3.3.1 Original Centers of Diversity of Brassica rapa

Being the first domesticated Brassica species, Brassica rapa has been cultivated for many centuries (over 4,000 years) from the highlands near the Mediterranean region to Scandinavia, to Germany and into Central Europe, and eventually to Central Asia (Gomez-Campo 1999; Dixon 2006). B. rapa was introduced into China through western Asia or Mongolia. The introduction into Japan could have occurred via China or Siberia. In Fig. 3-2, a map of Eurasia is depicted with the primary center and secondary centers of diversity for B. rapa are indicated.

A number of earlier studies based on morphology, geographic distribution, isozymes and molecular data indicate that B. rapa originated from two independent centers (Denford and Vaughan 1977; Song et al. 1988b; Gomez-Campo 1999). Europe is proposed as one center of origin.

Figure 3-2 A map of Eurasia is depicted with primary center and secondary centers of diversity for B. rapa indicated. Two possible routes are depicted for the introduction of B. rapa into China.
for turnip rape and turnip types, which were further developed in Russia, Central Asia and the Near East. Eastern Asia is proposed as another center of origin for Asian leafy vegetables. Other cultivar groups of *B. rapa* most likely originated from different morphotypes within the two centers of origin and subsequently evolved separately. Song et al. (1988b) considered that “sarson” and “toria” types in India were derived from the European turnip rape and have evolved separately. In studies of Zhao et al. (2005) and Warwick et al. (2008), the spring oilseed types including the yellow sarson types from India formed a subgroup separate from European and Asian groups, suggesting that the Indian subcontinent may represent the third center of origin and that the separate breeding tradition in this region led to the development of the sarson types.

Turnip is a very old *B. rapa* subspecies that was probably directly domesticated from the wild progenitor that arrived from the Iranian region into Europe (Reiner et al. 1995). In Europe, turnip has been cultivated since 2500–2000 BC and its cultivation spread to Asia after 1000 BC (De Candolle 1886). Many distinct types were known to the Romans at the beginning of the Christian era and some of those varieties bore Greek place names, indicating turnips were earlier cultured in the Roman Empire and Ancient Greece. Turnips were introduced to China and cultivated before Christ based on the Chinese book of poetry “Shih Ching” (Keng 1974). Nowadays Chinese turnips are often replaced by other vegetables and only consumed in the Xinjiang district. The broccoletto's originated from Italy, and form a clearly separate group somewhat related to European turnip and oil types (Zhao et al. 2005).

A large group of *B. rapa* is formed by the leafy vegetables differentiated into several subspecies or cultivar groups mainly from China and Japan. Chinese cabbage is native to China and two main hypotheses regarding the origin of Chinese cabbage exist: one is the hybridization hypothesis suggesting that Chinese cabbage originated from hybridization between turnip (or turnip rape) and pak choi (Li 1981). The loose-leaved type is the ancestral form and gradually developed into the heading form, which was selected for as an adaptation to cool temperatures. The other evolutionary hypothesis was proposed by Tan (1979), who suggested that Chinese cabbage was formed during the introduction of pak choi from southern to northern China. Cultivated forms of Chinese cabbage are mentioned later than pak choi in ancient Chinese literary records. During the Tang dynasty (AD 659), the loose-leaved Chinese cabbage was mentioned in the book of “Xin Xiu Ben Cao”. The pictures of semi-heading types appeared in the book of “Yin Shan Zheng Yao” in the Yuan dynasty (1330 AD). The heading Chinese cabbage originated between the Yuan and the Ming dynasty (1368–1644 AD), and became especially popular during the Ming dynasty. The origin of other Chinese leafy *B. rapa* types has been
discussed but is unclear up to now (Cao et al. 1997; Chen et al. 2000; Guo et al. 2002). Japanese vegetables are likely to be derived directly or indirectly from different types of ?pak choi, but have diverged through geographic isolation and intensive selection (Song et al. 1988b, 1990). In our experiment based on amplified fragment length polymorphism (AFLP) profiles (Zhao et al. 2005), Chinese shuicai accessions that resemble Japanese mizuna group in the ?pak choi cluster and do not group with the Japanese mizuna’s that form a distinct cluster.

It is believed that European forms and Asian types of oleiferous \textit{B. rapa} have different origins, involving the Mediterranean area and the region of central Asia, Afghanistan and the adjoining Indian subcontinent. There is a lot of evidence that European oilseed \textit{B. rapa} is genetically very close to the turnip type (Reiner et al. 1995). Domestication is believed to have occurred in the early middle ages in Europe. From the Indian oil crops of \textit{B. rapa}, brown sarson appears to be the ancestral type from which toria and yellow sarson types were selected (Gomez Campo 1999). The history of Chinese oleiferous \textit{B. rapa} domestication in China needs to be further clarified although evolutionary pathways have been proposed by some Chinese researchers, in which the common point is that it possibly derived from Chinese ?pak choi (Liu 1984; Cao et al. 1997). This is also supported by our result based on AFLP profiles, where Chinese oleiferous \textit{B. rapa} (Chinese turnip rape) clusters with ?pak choi (Zhao et al. 2005). The group of winter oil types from Pakistan is not directly related to either East Asian or European types based on AFLP analysis (Zhao et al. 2005). The STRUCTURE analysis based on allele frequencies confirmed that the Pakistan oil types shared genetic background with the Asian ?pak choi group and/or European turnip group (Zhao et al. 2007), which is possibly related to their breeding history. We considered that the winter oils from Pakistan may be an intermediate type that developed in Central Asia in a directional differentiation from the ancestral European type into leafy Asian types. Alternatively winter oils could be the ancestral type of the cultivated \textit{B. rapa} and have further developed into turnips in Europe and into pak choi in Asia. At present we are investigating which theory is correct.

\subsection*{3.3.2 Description of \textit{Brassica rapa} Crop Types}

\textit{Brassica rapa} is one of the agronomically important vegetable species and to a minor extent also an oilseed species in the genus \textit{Brassica}. \textit{B. rapa} vegetables are consumed worldwide and provide a large proportion of the daily food intake in many regions of the world. There is wide variation in the plant organs that are consumed, which has resulted in the selection of different morphotypes depending on local preferences. Based primarily on the organs used and secondly on their morphological appearance, a number of major
cultivar groups (Table 3-1; Fig. 3-1), which have been given subspecies names in the past, can be distinguished (Specht and Diederichsen 2001; http://www.plantnames.unimelb.edu.au/Sorting/Brassica_rapa.html).

Chinese cabbage: *B. rapa* L. subsp. *pekinesis* (Lour.) Hanelt

Chinese cabbage is native to China and is characterized by larger leaves and heads of different shape with winged petioles. It is mainly cultivated north of the Yangze River of China, in Korea and in Japan. The Chinese cabbage is used in autumn and winter for cooking or stewing dishes, but also for fresh salad and salt-pickled. In Korea, Chinese cabbage is used as the major component of “kim-chi”, the traditional preserved side dish and salad. At present the Chinese cabbage is commonly found in markets throughout the world.

Pak choi: *B. rapa* L. subsp. *chinensis* (L.) Hanelt

Pak choi does not form a head, is characterized by green-white, enlarged midribs. It is widely cultivated in southern and central China, and there is potential for the crop to be successfully grown in Southeastern Asia, including the Philippines, Malaysia, Indonesia and Thailand.

Wutacai: *B. rapa* L. subsp. *narinosa* (L.H. Bailey) Hanelt; *Brassica chinensis* L. var. *rosularis* Tsen & Lee

Wutacai (flat Chinese cabbage) forms a subgroup of pak choi-like cultivars that differ from typical pak choi types by their flat rosettes and many dark-green leaves. This crop is mainly cultivated in Southeastern China, and is more cold tolerant and resistant to bolting. It is used in cooked dishes and salads, but is also used for ornamentals.

Caixin (or Caitai): *B. rapa* L. var. *parachinensis* (L.H. Bailey) Hanelt

Caixin is an early flowering non-heading vegetable with similar leafy features to pak choi. It is mainly cultivated in southern and central China, and distributed in Southeastern Asian countries nowadays. The edible parts of this crop are the young inflorescences and stems that can be harvested 40–80 days after sowing.

Zicaitai: *B. rapa* L. var. *purpuraria* (L.H. Bailey) Kitam

Zicaitai is characterized by the purple red stem and non-heading phenotype, and is mainly cultivated in southern and central China. This flowering
purple-stemmed Chinese cabbage has tender early inflorescences, stems and shoots, which are edible. This vegetable is tolerant to low temperature and the purple color intensifies as the temperature decreases.

Taicai (or Tai tsai): *B. rapa* L. ssp. *chinensis* Makino var. *tai-tsai* Hort

Taicaís are non-heading cabbage cultivars with irregularly notched leaves of different blade shapes. The tender leaves, stems, and even the conical-shaped succulent taproots are edible. These types are mainly distributed throughout eastern China and are widely cultivated in the Shandong and Jiangsu provinces.

Mizuna and mibuna: *B. rapa* L. subsp. *nipposinica* (L.H. Bailey) Hanelt; *B. rapa* L. var. *japonica*

Mizuna and mibuna types are a small group of Japanese leafy vegetables with many serrated leaves or long narrow leaves. This crop is mainly cultivated and consumed in Japan. The shuicai cultivars from China resemble mizuna or mibuna, and Chinese fennie (tilling) vegetable with strong stooling leaves also belongs to this group (Cao et al. 1997).

Komatsuna: *B. rapa* L. subsp. *perviridis* Bailey

Komatsuna is a type of neep greens, which is consumed for its young leaves, stalks and flower shoots. It is mainly grown in Japan and is also known as Japanese mustard spinach. This vegetable can be stir-fried, pickled, boiled and added to soups or used fresh in salads. It is most often grown in the spring and autumn, as it cannot endure extreme heat or cold for more than a short time.

Turnip: *B. rapa* L. em. Metzg. subsp. *rapa*

The turnip types are a group of cultivars grown for their enlarged hypocotyl and taproot, which can be subdivided in vegetable and fodder turnips. Manifold shapes and colors are typical characteristics of turnips. The turnip preferably grows in misty and cold regions.

Broccoletto, Broccoli raab, Cima di rapa: Broccoletto group; *B. ruvo* L.H. Bailey

Broccoletto is a main Italian group of vegetable *B. rapa* of which the young compact inflorescences are consumed. Broccoletto has a strong stem and short internode length. The edible parts of this type are the small flower
heads that appear when the plants are about 20 cm tall. The edible part is quite similar to that of Chinese caixin, which is also utilized in the early flower stage.

**Turnip rape: B. rapa L. subsp. oleifera (DC.) Metzg**

Turnip rape is an oil type of *B. rapa*, which is mainly cultivated in Europe, China, India, Pakistan, Bangladesh and Canada, and there is potential for the crop to be successfully grown in the United States, South America and Australia. It falls into different subgroups based on its growth habit. Summer oil types (summer or spring turnip rape) are mainly cultivated in Canada, northern Europe and Bangladesh. Pakistan winter turnip rape is characterized by dissected leaves and rosette seedlings. Cultivars of winter oil type are still cultivated for oil and biomass production in Scandinavian countries, they are more cold tolerant than oilseed rape and have a high growth rate under low temperature. In China, three main oleiferous *B. rapa* ecotypes, viz. spring, winter and semi-winter turnip rape, were developed in adaptation to different climates, soil conditions, cultivation methods and farmer preferences (He et al. 2003), and have been grown as food oil and vegetable crop. Recently, the cultivated oleiferous *B. rapa* in both China and Canada was substantially replaced by the recently introduced *B. napus* cultivars, which have a higher yield and better adaptation.

**Sarsons: B. rapa L. subsp. dichotoma and trilocularis (Roxb.) Hanelt**

Indian oleiferous *B. rapa* includes three ecotypes, viz. brown sarson (*Dichotoma*), toria (*Dichotoma*) and yellow sarson (*trilocularis*). Brown sarson has long roots, with a limited lateral spread, enabling its successful cultivation under drier conditions. Toria has similar traits to brown sarson in morphology, and is believed to be selected from these types (Gomez-Campo 1999). The yellow sarson is characterized by very early flowering, self-compatibility and its yellow seeds.

**3.3.3 Phenotypic Diversity of Brassica rapa Vegetables and Oilseeds**

The *B. rapa* species is morphologically very diverse as mentioned above: genetic variation can be observed for many traits when different accessions are grown together under uniform environmental conditions. Within *B. rapa* species, a number of diversity studies have been conducted based on morphology, geographic distribution, isozymes and molecular data.

In the earlier researches on morphological diversity in *B. rapa*, which have been reviewed by Prakash and Hinata (1980) and Gomez-Campo
(1999), a number of cultivar subspecies was recognized and the three main types (leaf, turnip and oil) were classified. A study of morphological diversity within the turnip group (120 landraces) of a Northwestern Spain collection (Padilla et al. 2005), suggested that the classification of landraces provides valuable information about their characteristics and the variability between landraces could be used for breeding programs. In our recent experiment (Pino del Carpio et al. unpublished data), the hierarchical cluster analysis of 26 morphological traits from 163 accessions of *B. rapa* showed three distinct groups (1-the European turnip group, 2-the Chinese cabbage group, 3-the pak choi, brocoletto, Japanese turnip and winter turnip rape group) and two admixed groups. The further clustering analysis based on high-throughput profiling data of metabolic traits showed four main groups that relate to morphotypes and corresponded very well with the four groups that were defined using STRUCTURE based on AFLP and simple sequence repeat (SSR) markers. Phenotypic characteristics from accessions belonging to the different morphological groups can also be associated with classification based on genomic fingerprints, as was concluded from a study in our laboratory (Zhao et al. 2005). For example, leaf serration is a character that is associated strongly with the grouping in a phenetic tree. Turnips vary widely in shape and color, but as these characteristics are not associated with specific AFLP patterns they could not be differentiated between groups. However the European and Japanese turnips, which are different in leaf edge shape (European Turnips: dissected leaves; Japanese Turnips: no or mildly serrated leaves), also formed two distinct clusters based on AFLP data. The self-incompatibility trait also showed an interesting distribution in our experiment A (Fig. 3-3; Zhao et al. 2005). All of the pak choi of group 1 (PC1) and summer turnip rapes from group-Oil2 which includes yellow sarson, are self-compatible, while most pak choi accessions from group 2 (PC2), turnips from turnip-group 1 (T1), winter turnip rapes (Oil1), brocolletto (Bro), mizuna and komatsuna genotypes are self-incompatible. Incompatibility clearly differentiates subgroups that belong to the same cultivar groups as defined by similar use or phenotype, as it separates PC1 from PC2 and Oil1 from Oil2.

### 3.3.4 Genetic Diversity of *Brassica rapa* Vegetables and Oilseeds

In general, a high level of diversity was observed in the *B. rapa* gene pool based on isozyme and molecular data. Das et al. (1999) reported that 69% of 319 AFLP markers and 80% of 125 RAPD markers were polymorphic in nine accessions of *B. rapa*. McGrath and Quiros (1992) assessed 20 accessions including turnip, oil and vegetable types at isozyme and restriction fragment length polymorphism (RFLP) level, in which the highest level
Figure 3-3 UPGMA phenetic tree (left part) and structure image (right part) of 160 *B. rapa* accessions based on AFLP data. The phenetic tree indicates a loose population structure based on UPGMA using MEGA 2.1 software (Zhao et al. 2005), which shows two main subgroups (East Asia and Europe group) and a small subgroup (Sarson types from India). The population structure analysis is based on allele frequencies using the program STRUCTURE (Zhao et al. 2007), which confirmed the UPGMA result and suggested four subgroups. The structure picture was shown in accordance with the order of accessions in the phenetic tree. Accessions in structure are represented by a bar, which is partitioned into several segments with different gray shades according to the individual’s estimated membership fractions of the four clusters.
of heterozygosity was shown in central Asian and Indian types compared to European and Asian varieties. Genetic diversity among Nordic turnips comprising 31 germplasm collections, including varieties, elite stocks, landraces and older turnip slash-and-burn types, was studied using nine isozyme loci and 26 alleles (Persson et al. 2001). Genetic variation of 18.7% was observed within the used accessions, and the total heterozygosity value was 0.358. The level of heterozygosity was reduced in the landraces, but it was high within the variety group. In a recent study of Warwick et al. (2008), quite high diversity levels (62–79%) among 93 B. rapa accessions were revealed based on 307 AFLPs. Since the number and type of accessions, the evaluation system and the geographic origin are different for different reports, it is difficult to compare the genetic diversity levels. The AFLP fingerprints of a large collection of 160 accessions have shown that there is considerable genotypic variation in the B. rapa gene pool (Zhao et al. 2005). The amount of genetic variation present within groups of cultivars, as identified in the phenetic tree, contributed considerably to the high polymorphism rate observed. Low bootstrap values for many of the groups in the phenetic tree suggested that only a few genes are involved in causing the extreme morphologies. Within different cultivar groups, Chinese cabbage, pak choi, and turnip displayed high amounts of polymorphisms (75–82%), but two yellow sarson and two mizuna accessions had remarkably similar AFLP profiles, which may be caused by the limited number of accessions selected from those morphotypes. The high level of genetic diversity within B. rapa groups is possibly caused by their wide distribution in geographical regions with different growth habitats and consumer preferences. This could also explain why the different morphotypes could emerge independently in the different regions. Using SSR markers, Ofori et al. (2008) reported on the effect of crop improvement on genetic diversity in three turnip rape cultivars (32 plants per cultivar) representing three different breeding periods, and calculated that genetic diversity within cultivars accounted for 83% of total variation while between cultivar variation was 17%, suggesting that quality breeding did not significantly reduce the genetic diversity within cultivars.

3.3.4.1 Wild B. rapa and Its Relationship to the Different B. rapa Morphotypes

Wild B. rapa ssp. silvestris (L.) Janchen occurs mainly as a weed in arable fields and as a ruderal escape. It is regarded as the species from which cultivated turnip and turnip rape originated. It is difficult to distinguish wild silvestris from escaped cultivars of the same species (oleiforous B. rapa) on the basis of phenotype alone (Andersen et al. 2009). The relationship
between cultivated and wild \textit{B. rapa} was studied in an earlier RFLP study (Song et al. 1990), showing that the Chinese leafy types group was distinct from wild European populations and a wild \textit{B. rapa} accession from India was positioned between European and East Asian types. Crouch et al. (1995) reported the genetic diversity within five wild and seven cultivated \textit{B. rapa} accessions using 123 RFLP fragments and principal components analysis, suggesting that two wild Sicily and Algeria accessions were genetically distant from other forms, and two wild Argentina and California accessions were likely to be recent escapes from cultivation. A genetic diversity study with 131 inter-simple sequence repeat (ISSR) markers in 15 cultivated and 17 wild accessions of \textit{B. rapa} from Scandinavian countries was reported by Andersen et al. (2009): the similar genetic variation was observed between the two groups and diversity within wild populations was large (80%). The authors considered that most of the diversity in \textit{B. rapa} can be preserved by maintaining a rather small population with a representative geographical distribution.

3.3.4.2 Relationship of Genetic Diversity with Population Structure and Geographical Distribution

The genetic diversity within \textit{B. rapa} was also related to their wide distribution in geographical regions with different growth habitats and consumer preferences. In a study by Song et al. (1988b) based on RFLP markers, the cultivar types of \textit{B. rapa} were classified into a European turnip group including turnip and turnip rape and an East Asian group including leafy types. Based on genetic diversity analysis at isozyme and RFLP level, McGrath and Quiros (1992) concluded that crop types of \textit{B. rapa} used for similar purposes are derived from geographically differentiated populations. In a recent AFLP analysis of 32 \textit{B. rapa} accessions from the Eurasian continent (Takuno et al. 2007), the grouping of \textit{B. rapa} was also related to geographic origins. The authors suggested that cultivars in East Asia were probably derived from a cultivar group that originated from Europe or central Asia and migrated to East Asia. Cluster analysis in most of the diversity studies indicated that subspecies within \textit{B. rapa} showed a clear separation of two main phenetic groups: European turnip and turnip rape types; Asian turnip, leafy types and Chinese turnip rape. In the collection studied by Zhao et al. (2005), the oil type accessions from Canadian germplasm were missed. Those accessions were included in the recent diversity analysis by Warwick et al. (2008), and clearly were separated from the Indian sarson group and clustered with the European turnip group. Based on allelic frequency analyses (Zhao et al. 2007), it was also likely that part of the genetic background is shared by many accessions belonging to different crop types, suggesting that these accessions that
are genetically related possibly share part of their breeding history. These data provide insight into possible application of favorable alleles from accessions from the same morphotype and from different morphotypes of the gene pool for specific crop improvement programs, like introduction of resistance genes from different crop types. With knowledge of the allelic distribution, breeders can estimate the amount of backcrosses needed to retrieve the original crop type.

Leafy *B. rapa* vegetables distributed in East Asian countries show remarkable morphological diversity with local climatic zones. Turnips grown around the world can be classified into several groups by geographical distribution: Teltow turnips; West European turnips with dissected leaves; Asia Minor and Palestine turnips; Russian turnips; Asiatic turnips of Afghan type; Japanese turnips with entire leaves; and European entire-leaved turnips (Sinskaia 1928; Dixon 2006). Limited information is available on the genetic relationship between all these turnip types as measured by molecular marker screens in relation to the different geographical origins. Ren et al. (1995) reported that two Chinese turnip accessions were clustered in the Chinese cabbage group based on random amplified polymorphic DNA (RAPD) analysis among 30 *B. rapa* accessions. We assessed the genetic diversity in a large collection of *B. rapa* by AFLP fingerprinting, and the results showed that genetic similarities are more related to geographic origin (East Asia versus Europe) than to morphological groupings, suggesting either an independent origin and/or a long and separate domestication and breeding history in both regions (Fig. 3-3; Zhao et al. 2005). It was illustrated that European and Japanese turnips formed distinct groups, while Chinese turnips clustered with pak choi types (experiment B). It would be interesting to see whether Chinese turnips are closely related to the Chinese turnip rape. This geographical distinction of the turnips can also be seen in morphological and physiological characters such as leaf shape and flowering time and might either be due to a long separation of breeding of the different turnip types or even indicate an independent origin.

### 3.4 Diversity of *Brassica oleracea* Vegetables

#### 3.4.1 Origin of *Brassica oleracea*

Several genetic studies in *Brassica oleracea* looked into understanding the origin and the center of domestication of this *Brassica* species and are based on the relationship between the cultivated and the wild relatives. An indication of the close genetic relationship between cultivated and wild relatives is that all *n* = 9 *Brassica* species can be crossed easily and form hybrids and fertile offspring. Wild relatives of *B. oleracea* occur as perennial plants in rocky cliffs on the coast of the Mediterranean region.
and in northern Spain, western France and southern and Southwest Britain (Tsunoda 1980; Fig. 3-4).

The adaptations of these wild forms to different habitats have led to different cultivated forms as we know them today. The multiple origins of the different *oleracea* morphotypes can be derived from cross-breeding between closely related *Brassica* species growing in geographical proximity to each other. The introgressions from wild species have also been an important factor, which has increased the variability and adaptability of cultivated *B. oleracea*. Several theories have been drawn about the origin of the different forms, for example the origin of stem kales could be found in the *B. rupestris-incana* complex, characterized by the formation of a tall central stem and a large inflorescence before branching (Snogerup 1980). The origin of the bushy kales can probably be found in *B. cretica* given the common branching, shrubby habit and fleshy leaves. The white colored flowers may also indicate an origin derived from *B. alboglabra* and *B. cretica* subsp. *nivea* (Gómez-Campo 1999).

A significant aspect in *B. oleracea* is the development of cultivated forms in different countries for example broccoli, originated in Italy, diversified into different forms and Brussels sprouts developed near Belgium (Gómez-Campo 1999).

Cauliflowers are generally regarded as being derived from broccoli (Crisp 1982; Gray 1982). Crisp (1982) worked on hybridization between cauliflower and broccoli and identified one mutation in a gene in broccoli
as being responsible for the cauliflower phenotype. Purugganan et al. (2000) indicated that specific alleles of \textit{B. oleracea} \textit{CAULIFLOWER} gene (\textit{BoCAL}) were selected by early farmers during the domestication of modified inflorescence structures in \textit{B. oleracea}.

### 3.4.2 \textit{Brassica oleracea} Crop Types

The vast array of types developed within cultivated \textit{Brassica oleracea} has led to the acceptance of subspecies and variety (cultivar) descriptions based upon specialized morphology of edible parts. We based our description of groups on the classification made by Snogerup (1980) (Table 3.1).

**Brussels Sprouts: \textit{B. oleracea} L. var. gemmifera DC**

Biennial plants with simple erect stems and auxiliary buds, which later develop into compact miniature cabbage heads or sprouts. At the top of the stem there is a rosette of small leaves, which are generally petiolate and subcircular.

**Cauliflower: \textit{B. oleracea} L. var. botrytis L.**

Cauliflower is a biennial or annual herb, with unbranched stems; the leaves are arranged in a rosette, which surrounds the compact terminal flower head (curd). The curd consists of a dome of white to cream or yellow floral meristems that grow on short peduncles.

**Broccoli: \textit{B. oleracea} L. var. \textit{italica} Plenck**

Broccolis are heading types, which develop a large single terminal inflorescence. The main head consists of groups of fully differentiated green or purple flower buds, which are less densely arranged than cauliflower and with longer peduncles.

**Chinese kale: \textit{B. oleracea} L. var. \textit{alboglabra} (L.H.Bailey)**

Chinese kale is cultivated since ancient times without apparently obvious wild progenitors. It is an annual herb during the vegetative stage which reaches up to 1–2 m at the end of the reproductive stage. The leaves are crisp and thick.

**Other kales: \textit{B. oleracea} L. var. \textit{acephala} DC**

Many groups can be distinguished: curly kale, collard, marrow stem kale, palm tree kale, Portuguese kale and thousand headed kale. Kales are old
cultivated cole crops closely related and with morphological resemblance to the wild forms of *B. oleracea*.

**Kohlrabi:** *B. oleracea* L var. *gongylodes* L.

The Kohlrabi is a biennial crop in which the edible consumed part is the thickened stem.

**Chinesse cabbage:** White headed cabbage (*B. oleracea* L var. *capitata* L. *f. alba* DC), red headed cabbage *B. oleracea* L var. *capitata* L. *f.rubra* (L) Thell, savoy headed cabbage *B. oleracea* L. var. *sabauda* L.

All cabbages are characterized by the formation of tight compact heads, the head shapes vary from spherical, to flattened to conical. The leaves are smooth, curled or savoyed and the leaf color can be red-purple in red headed cabbage and green to yellow green in other cabbages.

### 3.4.3 Genetic Diversity of *Brassica oleracea*

Screening for genetic diversity has been widely used in *Brassica oleracea* to reveal the phylogeny of the different types (Astarini et al. 2004), to discuss relationships between populations (Geraci et al. 2001) and to analyze the diversification and origin of *Brassica* crops.

The genetic diversity studies in *B. oleracea* fall into different categories, some are focused in groups (wild and cultivated or both) and/or accessions from different regions and make use of different techniques like isozymes, RFLP, RAPD and DNA sequence analyses. Morphological analyses of different traits have been mostly directed to the preservation and characterization of landraces and cultivated types.

#### 3.4.3.1 Wild *B. oleracea* and *n* = 9 *Brassica* Species

As indicated above wild *oleracea* has been largely studied to compare population attributes. A large study with isozyme and RAPDs compared the population attributes among wild *B. oleracea* (Lanner-Herrera et al. 1996). From a germplasm of 44 wild populations of *B. oleracea* collected in Spain, France and Great Britain, 18 populations were selected for analysis.

The comparison of clustering with isozyme and RAPD of the same populations resulted in an isozyme based tree that displayed greater distances among populations than the RAPD based tree. In this study genetic distances did not appear to be related to geographic distances. The lack of clustering of populations from the same or proximal regions suggested an independent genetic differentiation of populations that occurred throughout the entire range of wild *B. oleracea*, a manifestation of the geographical isolation of wild *Brassica* populations.
Lanner-Herrera et al. (1997) analyzed the relationship between 10 wild Brassica species represented by 34 populations with RFLP markers. The UPGMA analysis showed that the species tended to cluster according to the geographic region: B. cretica and B. hilarionis comprise a cluster that could be called eastern Mediterranean; B. oleracea, B. bourgeaui, and B. montana define an Atlantic—western Mediterranean cluster; B. incana, B. rupestris, and B. villosa form an Italian group; and a B. insularis—B. macrocarpa association may be called central Mediterranean.

Other studies based on the classification of wild Brassica n = 9 taxa established the separation of Brassica species in groups from the West (B. oleracea, B. bourgeaui, B. incana, B. montana and B. alboglabra), central Mediterranean groups (B. rupestris, B. villosa, B. villosa subsp drepanensis and B. macrocarpa) and taxa that grow in the eastern Mediterranean (B. cretica and B. hilarionis)(Lazaro and Aguinagalde 1996, 1998).

Some of the studied species show a variable position in the clustering with others. For example, B. incana, which belongs to the rupestris group appeared in a study of Lazaro and Aguinagalde (1996) more closely related to B. oleracea, probably because of a former introgression by human influence. In B. cretica, the considerable morphological differentiation and events like genetic drift and founder effects can explain the extreme variation between isolated populations and differential relationship with other wild species.

Lanner (1998) also focused on the study of a specific small, variable region in the chloroplast sequence to compare 34 populations representing Brassica bourgenaui, B. cretica, B. hilarionis, B. incana, B. insularis, B. macrocarpa, B. montana, B. oleracea, B. rupestris and B. villosa (Lanner 1998). This study investigated the relationships that may be detected in a phylogenetic analysis compared with previous phenetic analyses of Lanner-Herrera et al. (1997) and Song et al. (1990). Three groups are well distinguished: one including B. bourgenaui, B. oleracea, B. cretica, B. incana, B. hilarionis, B. insularis and B. montana, which subdivides into Brassica bourgenaui-B. oleracea and B. cretica-B. incana clades. A second group includes B. insularis and B. macrocarpa and a third group comprises B. rupestris and B. villosa.

Several conclusions were drawn with respect to the relationship between the wild Brassicas. The majority of populations of the same species had the same chloroplast haplotype, with divergence observed in two species B. montana and B. cretica. B. insularis also showed variability, which demonstrated the distinction between populations from Tunisia and those from Sardinia and Corsica. The presence of B. macrocarpa genetic material in a population of B. insularis was considered as evidence of introgression from B. macrocarpa into B. insularis.

The comparison with previous RFLP studies (Song et al. 1990; Lanner-Herrera et al 1997) agrees and provides additional evidence of the
close relationship between B. villosa and B. rupestris and a more distant relationship to B. macrocarpa separated in a different clade.

In the case of the haplotype similarity between B. cretica and B. incana, the results differ from RFLP studies, which showed that B. incana was associated with B. rupestris and B. villosa. The consensus indicates that the wild species have close relationships and have not diverged significantly since their separation from a common ancestor.

The most divergent species B. rupestris, B. villosa and B. macrocarpa were also found as a different group with isozymes (Lazaro and Aguinagalde 1998) and this study identifies the Sicilian region as a center of genetic diversity.

The study of diversity with SSRs in the chloroplast genome, including domesticated and wild B. oleracea and nine inter-fertile related wild species, was performed by Allender et al. (2007). The results corroborate to the high level of diversity in B. cretica with the presence of multiple chloroplast haplotypes as shown already in the study of Lanner et al. (1996). However, the lack of diversity found within B. oleracea was remarkable in the study of chloroplast sequences. Panda et al. (2003) also found no variation in chloroplast PCR-RFLP profiles between a cultivated accession and nine wild populations from Spain, France and the UK.

The lack of diversity in this study pointed towards a single center of domestication in the eastern Mediterranean region (Gomez-Campo 1999) for B. oleracea.

3.4.3.2 Cultivated Brassica oleracea

Several examples can be found for the study of genetic diversity within types of cultivated B. oleracea. Allen et al. (1986) searched for markers at the enzyme level to distinguish the origin of cauliflower stocks. The investigation was restricted to cauliflowers available in the United Kingdom, and included annuals and biennials together with Australian types. The isozymes examined showed evolutionary divergence of the cauliflower types during the selection for different times of development. The results also revealed that European annuals and biennials are parents of the Australia cauliflowers.

The usefulness of different marker systems to examine genetic variation within types has also been compared in several studies. The genetic variability of kale plants (Brassica oleracea L. var. acephala D.C.) was studied by means of enzymatic polymorphism using polyacrilamide gel electrophoresis and a DNA polymorphism assay based on RAPD (Sawaza et al. 1997). Fifteen clones of kale B. oleracea var. acephala were studied and differences among clones were observed by isoenzymes and RAPD. However, the dendrogram obtained from both markers were dissimilar, suggesting that the isozymes provided less information than the RAPD about the genome diversity.
A comparison between two marker systems was done in the study of wild cabbage in Dorset (UK), where Raybould et al. (1999) investigated if the restricted pollen and seed flow was the cause of population substructure and isolation by distance. Ten populations were screened with microsatellite primers designed for oilseed rape and compared with data of isozyme loci. Isozymes were highly polymorphic within the population, very similar to the results from British populations investigated by Lanner-Herrera et al. (1996) with isozymes studying wild accessions collected in Spain, France and the United Kingdom. The oilseed rape SSR primers were effective at detecting genetic variation but the number of alleles detected was low with these specific primers. The results showed Dorset populations were isolated by distance with restricted but with significant gene flow among them in Dorset.

Several *B. oleracea* cultivars are restricted to localized areas and/or countries, and these have been subjected to the evaluation of genetic diversity. One case study is on the endangered Italian cultivar Broccoli Fiolaro, which is a local broccoli variety restricted to an area in Northeastern Italy. The study by Vischi et al. (2008) highlights the genetic distinction of the Fiolaro cultivar from other Italian cultivars and the distribution of polymorphism in a geographical pattern. Other studies within types combined morphological data with molecular marker results. In the study of a Turkish kale population the fact that populations clustered according to geographical distribution but not to their morphological characteristics was the most significant result (Okumus and Balkaya 2007).

In a broader study, Lamboy et al. (1994) evaluated the genetic diversity among 56 different *B. oleracea* accessions comprising 14 botanical varieties. The results present that isozymes are partially useful because specific alleles were not limited to a group of accessions but were dispersed all throughout the morphological groups. In this study, the use of RAPD and RFLP markers was suggested for the identification of accessions or botanical varieties. However, Van Hintum et al. (1996) evaluated a selection of 11 white cabbages and nine Brussels sprouts using isozymes, and could classify the groups as expected; all the Brussels sprouts clustered together as did the white cabbages.

Broccoli and cauliflower have been found to separate into different groups with RAPD markers (Hu and Quiros 1991). In the same study, the level of polymorphism detected in cauliflower cultivars was significantly higher than in broccoli. However, when diversity had been evaluated based on database-derived SSR markers, it was found to be low in cauliflower, intermediate in broccoli (Calabrese) and greatest in cabbage (Tonguç and Griffiths 2004).

In an RFLP study, Song et al. (1988a, b) revealed that *B. oleracea* accessions could be divided into three groups: a cabbage group including collard, kohlrabi, Portuguese cabbage, borecole and Savoy cabbage; a
broccoli group including marrow stem, Jersey kale and Brussels sprouts; and a group including thousand head kale.

The differential grouping of broccoli, cabbage and in particular the distinctness of cauliflower is in accordance with the results from Nienhuis et al. (1993) who also grouped these botanical varieties into different clusters based on RFLP polymorphisms.

Portuguese tronchuda cabbages and kales, two peculiar vegetable types, are of genetic interest since they are only located in particular regions, and can only be found in Portugal or in regions with Portuguese influence. Extensive studies on landraces of tronchuda cabbages and kales indicated similarity based on RFLP patterns, resulting in tronchuda and kales clustering within the same group (Dias et al. 1991). In a study of tronchuda cabbages and kales, which included several other B. oleracea types, the clustering tree diverges in groups including broccoli and cauliflower, a group with a mixture of kales and cabbages originally from central-North Europe, and one group formed by Portuguese coles. The results of these studies suggest the existence of three major regions of domestication of B. oleracea in Europe: Italy, central-North Europe and Portugal (Dias 1995).

3.4.4 Phenotypic Diversity of Brassica oleracea

Based on the morphological appearance of different varieties, Brassica oleracea has been statistically classified into groups.

Dias et al. (1993) conducted a study using morphological characters to determine relationships among Portuguese cole landraces using 58 accessions belonging to Portuguese tronchuda cabbage (B. oleracea var. tronchuda Bailey syn. Var. costata De Candolle, Couve Tronchuda), Portuguese galega kale (B. oleracea var. acephala De Candolle, Couve Galega) and other coles such as Algave cabbages (B. oleracea var. capitata L., Couve do Algave or B. oleracea var. sabauda L., Couve Repolho-lombarda do Algave). The accessions were evaluated on 46 morphological characters from seedling stage to ripe silique under field conditions and in two consecutive years. After analysis using clustering by UPGMA, eight main groups were found. Five groups corresponded to tronchuda cabbage landraces and three groups corresponded to cabbages, miscellaneous coles and galega kales. Groups of landraces were primarily associated with morphological differences among accessions and secondly with geographical origin.

Cartea et al. (2003) evaluated the morphological and agronomic characteristics of 15 kale populations. The aim was to determine relationships and also identify morphological characteristics that could be incorporated into breeding programs. The cluster analysis resulted in the identification of five groups, two groups including most of the populations and three small independent groups. Groups of landraces were associated
primarily with the geographical origin and secondarily with morphological differences among populations. Clusters separated according to the length of their vegetative phase, North and early populations (Cluster A) and South and late populations (Cluster B). The results of this work are in disagreement with the morphological classification of landraces of *B. oleracea* from Portugal performed by Dias et al. (1993). Although it should be noted that Dias et al. (1991) found that with molecular data (RFLPs) in the same collection the clusters were defined on the basis of their geographical origin and not on the basis of their morphological differences.

As reported above, morphological cluster analysis within crop types has been related often with geographical origin. Balkaya et al. (2005) studied similarities and differences regarding morphological variation of 95 white head cabbage populations as genetic resources collected from different eco-geographical regions of Turkey. The study reported the existence of considerable morphological variability. However, cluster groups were not associated with the geographical origin of the genotypes.

The other important aspect in phenotypic classification is the identification of variability in the adaptation of *B. oleracea* cultivars and landraces to different seasons. For example 21 landraces of cabbage (*B. oleracea* var. *capitata* L.) and two of tronchuda cabbage (*B. oleracea* var. *tronchuda* Bailey) from Galicia (Northwestern Spain) along with five commercial hybrids of cabbage and one commercial variety of tronchuda were evaluated in two planting dates for an early and late harvest (Padilla et al. 2007). The results showed that head forming landraces seemed better adapted to the autumn/winter season than commercial hybrids, with higher yields, better early vigor, and less plants lost after transplanting. However, for characters related to the uniformity of the crops (appearance, days to harvest and synchrony of production), commercial hybrids available in local markets did better than landraces.

### 3.5 Diversity of *Brassica juncea* Vegetable and Oil Mustards

#### 3.5.1 Origin of *Brassica juncea*

*Brassica juncea* is a predominantly self-pollinated allopolyploid species with the A genome of *B. rapa* and the B genome of *B. nigra* and is widely grown as an oilseed in India and Pakistan, as a vegetable in China and Japan, as a condiment in North America and Europe, and more recently as Canola quality oil in Canada and Australia. The cultivation history in China could date back to the 6th century (Liu 1996), while Gomez-Campo (1999) even mentioned records of Su’s work (10–61AD) “Tu-Jin-BinCao” a mustard that was a popular crop at the time. Its economic importance and
the enormous morphological variability have attracted many researchers, including botanists, taxonomists and breeders since the 17th century. Several studies of comparative morphology, biochemical and molecular diversity provide evidence that the origin of *B. juncea* is not monophyletic but polyphyletic (Vaughan 1977; Prakash and Hinata 1980; Song et al. 1988a; Jain et al. 1994). The region of the Middle East is a likely center of origin because of the occurrence of the wild forms of *B. rapa* and *B. nigra* (Gomez-Campo 1999). Two secondary centers of variation are proposed. North West India harbors the oilseed mustard, and Southwest China harbors the vegetable types (Dixon 2006) (Fig. 3-5). Parallel variation can be observed between *B. rapa* and *B. juncea* in East Asian vegetable forms like the loose head savoy, heading types, stalked types and swollen root types (see Table 3.1), which suggests independent interspecific hybridizations between *B. rapa* types and *B. nigra* that resulted in the different *B. juncea* types (Song et al. 1988a). In *B. juncea* the B-chromosomes show homeologous regions derived from *B. nigra* while two of the three *B. rapa* homologous segments for an 8.7 Mb *Arabidopsis* chromosome 4 segment show an altered structure as observed using GISH (Lysak et al. 2005). Earlier Axelsson et al. (2000) concluded that parental genomes were stably maintained based on an RFLP study. Similarly Panjabi et al. (2008) used intron polymorphism markers and concluded that A and B genomes of *B. juncea* were highly collinear with

---

**Figure 3-5** A map of Eurasia is depicted with primary center and secondary centers of diversity for *B. juncea* indicated. In India the mustard oils evolved, while in China the vegetable mustards evolved. At present *B. juncea* is widely grown as seed oil in India and Pakistan, as a vegetable in China and Japan, as a condiment in North America and Europe and recently as Canola quality oil in Canada and Australia.
A and B genomes of *B. napus* and *B. nigra*; comparison to the A genome of *B. rapa* was however not made. In a study based on restriction patterns of the chloroplast genomes in *Brassica* amphidiploids, evidence was presented that *B. juncea* has the chloroplast genome identical to *B. rapa* (Erickson et al. 1983). In another study, 40 chloroplast mutations that were phylogenetically informative were used to construct a parsimonious tree that showed that the *B. juncea* chloroplast genome was similar to the *B. rapa* genome (Palmer et al. 1983). Since only a few accessions were studied in both reports, this conclusion needs to be viewed with caution since *B. juncea* could have originated from several independent hybridization events between the parental species (Song et al. 1988a).

### 3.5.2 Description of *Brassica juncea* Crop Types

*Brassica juncea* is an erect annual to biannual herb, normally unbranched, sometimes with an enlarged taproot. Rosette leaves are very variable in shape and size, some curved and overlapping, some open and spreading, leading to heading and non-heading types. Leaves are entire, divided, with thin or thick petioles and with flat or tumor like surfaces. The fast array of types developed within the cultivated *B. juncea* has led to the acceptance of subspecies and variety (cultivar) descriptions based upon specialized morphology of edible parts. Nomenclature is rather complex, since different authors use different subspecies names. Here nomenclature as described by Gladis and Hammer (1992) is followed, with sometimes references to other nomenclatures.

**Leaf Mustards** (*B. juncea var. rugosa* Bailey; *var. latipa*, *var. crispifolia*, *var. foliosa*, *var. leucanthus*, *var. linearifolia* and *var. multisecta*)

Large plants have broad, large, green or purple leaves (*rugosa* and *latipa*) and plants have dissected leaves (*var. crispifolia*, *var. foliosa*, *var. leucanthus*, *var. linearifolia* and *var. multisecta*).

**Head and Curled Leaf Mustards** (*B. juncea var. capitata* and *var. involuta*)

Plants have leaves with broad petioles and midribs; leaves overlap or form a globose head.
Strumiferous Mustards (B. juncea var. strumata)
Plants have large leaves and well developed petioles with cushion-like perturbances at the top.

Tillering Mustards (B. juncea var. gemmifera and var. multiceps)
Plants have many lateral buds on a shortened stem which grow as tillers before flowering.

Swollen Stem Mustards (B. juncea var. tumida; also named B. juncea var. tsatsai Mao)
Plants have swollen stems with tumor like perturbances just below the petioles, with large leaves that are entire or dissected and red or green.

Swollen shoot mustards (B. juncea var. crassicaulis)
Plants have oblong or ovate entire or dissected leaves and shoots are swollen. This crop is distributed in Mongolia, Manchuria and northern China and acclimated to the cold unlike other B. juncea types.

Root Mustard (B. juncea var. megarrhiza also called B. juncea var. napiformis Paill. et Bois)
Plants have fleshy cylindrical or conical taproots with purple or green, entire or dissected leaves.

Oilseed Mustard (B. juncea var. juncea also named B. juncea var. gracilis Tsen et Lee)
This crop with pinnate leaves is distributed in India, central Asia and Europe, with the majority cultured for oil, even though greens are eaten from young plants.

Multi Seed Stalk Mustard (B. juncea var. utilis Li)
This crop with pinnate leaves is distributed in India, central Asia and Europe, with the majority cultured for oil, even though greens are eaten from young plants.
3.5.3 Genetic Diversity of Brassica juncea

Many studies have been conducted that assess the genetic diversity and phenotypic variation among Brassica juncea germplasm, however the choice of germplasm with regard to type (oil, mustard or vegetable), geographical origin (India, Pakistan, China, Canada and Australia) and numbers, hamper direct comparisons between these studies. Another complication is the nomenclature of the different types, since different authors use different subspecies names.

Fifty-two mustard (B. juncea) accessions, mainly Pakistani landrace mustard types plus a few vegetable cultivars from Japan and a few oilseed cultivars from India and Pakistan were grown in the field in Japan and characterized for 35 agro-morphological characteristics from seedling emergence to maturity (Rabbani et al. 1998b). Since morphological variation within the Pakistani landraces and oil types was low, while variation between these groups and the Japanese vegetable types was high, the authors concluded that mustard in Pakistan has a narrow genetic base, which was confirmed with genetic studies based on DNA polymorphisms and based on SDS-PAGE analysis of total seed proteins (Khan et al. 2008; Rabbani et al. 1998c, 2001). The genetic diversity in a similar collection of mustard (B. juncea) that included 41 landraces from Pakistan, six oilseed cultivars or lines from Pakistan, India, China and Australia and five Japanese vegetable cultivars was very low for the oilseed types as measured using 198 polymorphic RAPD fragments, indicating high genetic similarity (Rabbani et al. 1998c). Variability within the cluster of Japanese vegetable mustard was much higher than that of Pakistani accessions. A dendrogram constructed using UPGMA illustrated that Pakistani accessions and the six oilseed cultivars did not form distinct clusters similar to the morphological study. A similar conclusion was drawn in another study of genetic diversity among Indian mustard genotypes from Pakistan comprising 37 germplasm collections, five advanced breeding lines and three improved cultivars based on RAPD markers (Khan et al. 2008). The authors concluded that this low level of genetic diversity was due to intensive selection pressure for uniform phenotypes or due to a narrow genetic base of Pakistani germplasm (Khan et al. 2008). Jain et al. (1994) also used 595 RAPD markers to study the genetic diversity and genetic relationships in Indian mustard (B. juncea) among 12 Indian and 11 exotic B. juncea genotypes from Russia, Australia, and each one from Poland, the UK, Pakistan, and the USA. A dendrogram constructed using UPGMA cluster analysis of Jaccard genetic similarity coefficients revealed two groups: group A contained seven exotic accessions from Russia, Australia and Poland, while group B comprised the 12 Indian accessions in one subcluster and four exotic accessions that were in a distinct subcluster. The pedigree of Russian and Australian germplasm of group
A was not described; however some Australian accessions had Russian names, suggesting a common origin. The authors state that grouping of the genotypes according to geographic diversity is understandable given the polyphyletic rather than monophyletic origin of *B. juncea*. Another interesting result from the study of Jain et al. (1994) is that they recorded high percentages of heterosis in crosses between Indian and exotic genotypes, while crosses within Indian or exotic genotypes often exhibited negative heterosis. In yet another study, AFLPs were used to assess the genetic diversity among 21 established natural and nine synthetic varieties and lines of oilseed *B. juncea* from Asia, Australia, Canada, Eastern Europe and Russia to identify new sources of variation for hybrid breeding (Srivastava et al. 2001). Clustering using the UPGMA method based on 778 polymorphic out of 1,251 total fragments resulted in three distinct clusters. The Indian and Chinese genotypes plus previously developed synthetics formed one cluster, the newly developed synthetics formed yet another cluster and the lines from Australia, Canada, Eastern Europe and Russia formed the third cluster. Genetic diversity and the relationship of 47 local mustard accessions from Sri Lanka were estimated using 589 fluorescent amplified fragment length polymorphism (FAFLP) markers and several agro-morphological characters (Weerakoon et al. 2008). FAFLP and agro-morphological based clusters were in poor agreement, and resulted in grouping of the 47 accessions into different numbers of clusters. The FAFLP analysis seemed more accurate since it clearly separated the six canola varieties into a single cluster, while with agro-morphological analysis they were scattered among mustard accessions (Weerakoon et al. 2008).

The genetic diversity of selected breeding lines and cultivars of canola quality *B. juncea* was also assessed using AFLPs (Burton et al. 2004). Seventy-seven lines from three of the world’s major breeding programs, two from Canada and one from Australia, plus 15 mustard quality *B. juncea* accessions from India, China, Russia and Australia were included in this study, which resulted in the partitioning into five main groups; the mustards formed two groups, Indian mustards and the mustards from the other origins, while the canola quality accessions formed three groups, two of which corresponded to two breeding programs, while the third included lines from all three breeding programs. Variation within breeding programs was larger than variation between them and implications for further breeding strategies were discussed.

A number of studies applied different marker types (RAPD, ITS, AFLP) to reveal genetic relationships among small sets of vegetable mustards (Ren et al. 1995; Qiao et al. 1998; Fu et al. 2006; Qi et al. 2007, 2008). In all these studies, the accessions formed separate clusters but often the different vegetable types like root, expanded stem, leaf and seed stalk mustards grouped together and did not form separate groups. Thus phenotypically
similar accessions often possess different genetic backgrounds, suggesting that the diverse morphotypes were selected multiple times and that high selection pressure during domestication resulted in abundant morphological variation in genotypes with similar genetic backgrounds (Qi et al. 2008). No obvious grouping was detected according to geographical origin. The above described studies focus on either oil, mustard or vegetable B. juncea; it would be very interesting to compare diversity between Indian and exotic mustard oilseeds on the one hand with Asian vegetable mustards and these mustard oilseeds on the other hand. This comparison was made in a study of Wu et al. (2009), where 95 oilseed and vegetable mustard accessions from China, France, India, Pakistan and Japan were assessed for genetic diversity using 326 scorables fragments (with 161 polymorphic) generated using SRAP markers. They clearly demonstrated that genetic diversity within vegetable mustards was much higher than within oilseed mustards, and also that winter oilseed mustards are genetically more diverse than spring oilseed mustards. The vegetable mustards, winter oilseed mustards and spring oilseed mustards formed three distinct groups, with spring and winter oilseed mustards genetically closer to each other compared to the vegetable mustards, even though the vegetable and winter oilseed mustards share the same geographical origin (China, and even provinces within China). Based on these data they also conclude, like Song et al. (1988a), that oilseed and vegetable types evolved independently and further deduce that based on the higher level of genetic diversity, vegetable types were domesticated earlier than oilseed-types. Unlike the obvious grouping of oilseeds and vegetable mustards, based on the organs used, did not cluster together, similar to results of other studies (Ren et al. 1995; Qiao et al. 1998; Fu et al. 2006; Qi et al. 2007, 2008). This suggests that selection for different forms affected only a few genes that directed the morphological changes in plant organs. Similar results were found in a study of genetic diversity of the vegetable B. rapa types (Zhao et al. 2005).

3.6 Diversity of Raphanus sativus

3.6.1 Taxonomic Relationship of Vegetable Brassicas with Raphanus sativus

Phylogenetic analyses in the subtribe Brassiniae based on RFLPs of chloroplast DNA, indicated a clear division of this subtribe into two main lineages that were designated as rapa/oleracea and nigra (Warwick and Black 1993, 1997). The genus Brassica was polyphyletic as species within this genus were placed in two groups and Raphanus was placed in the rapa/oleracea lineage. In another study, sequence analysis of two chloroplast noncoding regions from seven species also showed that R. sativus was more related to the rapa/oleracea lineage than to the nigra lineage (Yang et al. 2002). Lu
et al. (2008) analyzed nucleotide sequence variations in the mitochondrial trnK/MatK region of cultivated and wild radish and related *Brassica* species and also concluded that *Raphanus* accessions formed one cluster that was located in the *rapa/oleracea* lineage, which had diverted from the *nigra* lineage. They further mentioned that *Brassica barreleiri* was the phylogenetically closest *Brassica* species to *Raphanus*. Flannery et al. (2006) developed plastid SSR markers based on 11 gene regions and characterized the plastid genome in several *Brassica* species and Brassicaceae using eight of these plastid SRRs. The analyses failed to separate *B. rapa* and *B. napus* individuals in clearly identifiable groups, but identified a monophyletic group that contained *B. rapa*, *B. oleracea*, *B. napus* and *B. nigra*, with a sister group *Raphanus*. These results did not support the hypothesis that *Raphanus* is more related to a *rapa/oleracea* lineage than to a *nigra* lineage (as suggested by Warwick and Black 1993, 1997 and Yang et al. 1998). Many comparative mapping studies suggest that the *Brassica* genomes descent from a common hexaploid ancestor, and are variants of three rearranged ancestral genomes (Parkin et al. 2005; Schranz et al. 2007). A study of comparative chromosome painting using *Arabidopsis* BAC contigs that cover 8.7 Mb of chromosome 4 (At4-b contig) to meiotic chromosomes of 21 species in the Brassicaceae to look for genome triplication and compare that with phylogenetic data obtained from chloroplast sequences (5′-trnL-trnF region), concluded that taxa with triplicated At4-b were in the single major Brassicaceae clade with three groups (Lysak et al. 2005). The largest group contained the *rapa/oleracea* lineage with *Raphanus*, and the *nigra* lineage. In none of the three-copy species, the same structural rearrangements were found. Song et al. (1988a, 1990) studied phylogeny of *Brassica* and related genera based on nuclear RFLPs. They presented evidence for the existence of two basic evolutionary pathways for the diploid species: one pathway gave rise to *B. nigra*, *B. fructulosa*, *Sinapis arvensis* and more distantly to *R. sativus*, the other gave rise to *B. oleracea* and *B. rapa*. *R. sativus* represented an intermediate and Song et al. (1990) postulated that it might have been derived from introgression or hybridization between species belonging to the different lineages. Molecular phylogenetic studies based on the internal transcribed spacer (ITS) region of 18S–25S rDNA of *Brassica*, *Rorippa*, *Arabidopsis* and allied genera showed that within the tribe Brassiceae, *Raphanus* is more closely related to *B. nigra* than to the *B. oleracea/B. rapa* clade (Yang et al. 1999). Also Thorlmann et al. (1994) used both RFLP markers, detected by genomic DNA and cDNA, and RAPD markers to evaluate the genetic relationships among 18 accessions from six cultivated *Brassica* species and one accession from *R. sativus*. Based on the RFLP results, *B. niger* was placed closer to *R. sativus* than to any other *Brassica* species, while based on the RAPD analyses, *B. niger* was closer to *B. juncea* and *B. carinata*, but the position of *R. sativus* was still closer to *B. niger* than to the group with *B. rapa*, *B. napus* and *B. oleracea*. These results
are in agreement with the nuclear data obtained in several studies (Song et al. 1988a, 1990), but are in conflict with the analyses based on mitochondrial and chloroplast DNA. To explain this discrepancy, the authors postulated the possibility that *Raphanus* is a hybrid between the *B. nigra* and *B. oleracea/B. rapa* lineages with the latter as the maternal parent.

### 3.6.2 Origin of *Raphanus sativus*

*Raphanus sativus* (*2n = 18*) is an old and globally cultivated oilseed, vegetable and forage crop, which is morphologically extremely variable. As center of domestication of the species, the East Mediterranean region and adjacent Near East are often mentioned (reference book Mansfeld Encyclopedia) (Fig. 3-6). The radish genus includes besides the cultivated species several wild species, like *R. raphanistrum*, *R. landra*, and *R. martimus*, all distributed in the coastal area of the Mediterranean Sea. *R. sativus* var. *raphanistroides* is also a wild species, which is distributed in the Far East, mainly in coastal areas of Japan, Korea and China. It is not clear what the ancestor of the cultivated radish is, even though *R. raphanistrum* is often mentioned.

![Figure 3-6](image-url) A map of Eurasia is depicted with primary center and secondary centers of diversity for *R. sativus* indicated. *R. raphanistrum*, *R. landra* and *R. martimus* are distributed in coastal areas of the Mediterranean Sea, while *R. sativus* var. *raphanistroides* is distributed in coastal areas of Japan, Korea and China.
3.6.3 Description of Raphanus sativus Crop Types

The many different crop types within cultivated radish have led to the acceptance of subspecies and variety (cultivar) descriptions based upon specialized morphology of edible parts. Nomenclature is rather complex, since different authors use different subspecies names. Here the three main radish types are described; for nomenclature we refer to the following website: [http://www.plantnames.unimelb.edu.au/Sorting/Raphanus.html](http://www.plantnames.unimelb.edu.au/Sorting/Raphanus.html).

**Oil radish, R. sativus convar. oleifer (Stokes) Alef.**

The radish oil crop has been cultivated since ancient times in East Asia, Egypt and perhaps also in India and since the 19th century in some European countries as well. The oil was used for food but also as lamp oil. In West and central Europe, the crop gained importance as green forage, silage and green manure plant. In Japan and Egypt cultivars are grown that are cultivated for their edible leaves, and are considered belonging to this group.

**Serpent, Seedpod or Rat-tailed Radish; R. sativus convar. caudatus (L.f.)**

This crop is cultivated in Malaysia, Indonesia, Sri Lanka and India, mainly along the western coast of Punjab, for their extremely long young siliques, which are eaten fresh or in pickles, or sometimes as a cooked vegetable.

**Radish, R. sativus convar. sativus L.**

This is an assemblage of different cultivar groups of both radish and small radishes, cultivated worldwide. The two main groups are the radish group (also named *R. sativus* var. *hortensis*), and the small radish group (also named *R. sativus* var. *sativus*). There is enormous variation in shape, size, and color of the tubers, leaves and siliques but also in seasonal development with hundreds of local landraces and cultivars. Radishes are used fresh in salads, but also in cooked or stewed dishes and salted dried or fermented. Small radishes are mainly used as a fresh vegetable, and are also variable, but less compared to the true radishes. Some reports also list *R. sativus* var. *niger* (black radish).
3.6.4 Genetic Diversity of *Raphanus sativus*

3.6.4.1 Cultivated Radish

Genetic diversity of 30 radish accessions from Pakistan was investigated both at the phenotypic and genotypic level with morphological characteristics and also at the DNA level using RAPDs (Rabbani et al. 1998a). Dendrograms were generated based on the morphological data and 202 RAPD bands. Based on morphological traits, the accessions were classified into four major groups that correspond to the four different forms of cultivated radish, large rooted, pod-type and oilseed-seed type. However, based on RAPDs, the accessions were grouped according to their collection sites and not based on their phenotypic differences. A number of other studies in both cultivated and wild radish also showed close relationships between genetic distances and geographical distances (Huh and Ohnishi 2001, 2002; Liu et al. 2008; Wang et al. 2008). Thirty-five late bolting radish cultivars were fingerprinted with RAPD, ISSR and SRAP markers, which resulted in similar estimates of polymorphism rate and genetic similarity coefficients between pair of genotypes (Liu et al. 2008). Clustering using UPGMA showed clusters that were in accordance with geographical origins (Korea, Japan and China) and main characteristics. In another study, 65 accessions of cultivated radish collected from 21 Eurasian and African countries were studied using AFLPs (Wang et al. 2008). In a neighbor-joining tree, the accessions formed groupings according to their geographical origin (Europe, Middle-East, South Asia and East Asia). The branching pattern in the neighbor-joining tree provides some information on the distribution route followed by radish, from the Mediterranean to East Asia via the Middle East and Central Asia or South Asia, similar to the introduction from the Mediterranean area along the silk route to China as suggested by Kitamura (1958) and Kumazawa (1963). Sixty eight European cultivated radish varieties, including open-pollinated varieties, inbred lines, 2n and 4n hybrid varieties of garden radish (*R. sativus var. sativus* DC. Convar. *radicula*) and black radish (*R. sativus var. niger*) were fingerprinted using AFLP and ISSR markers. Cluster analyses separated the two weedy species (*R. raphanistrum*) from the cultivated varieties, while within the cultivated germplasm black radish and French breakfast radish types with elongated roots formed separate clusters (Muminovic et al. 2005).

3.6.4.2 Wild and Weedy Radish and Their Relationship to Cultivated Radish

To clarify the relationship between wild and cultivated radish species (both European *R. landra* and *R. raphanistrum* and the Far Eastern *R. sativus*)
var. *raphanistroides*) and to identify the ancestral species of radish, nine accessions from three wild species and nine accessions of different cultivated varieties together with closely related *Brassica* species were investigated for differences in the *trnK/matK* sequence (Lu et al. 2008). Both wild and cultivated *Raphanus* accessions formed a cluster and belonged to the *rapa/oleracea* group, while the *nigra* group was more distantly related, similar to the findings of Warwick and Black (1993, 1997) and Warwick and Sauder (2005). Multiple origins were suggested for the cultivated radish, since each of the three variety types—European small radish *R. sativus* var. *sativus*, Spanish Black radish *R. sativus* var. *niger* and East Asian big long radish *R. sativus* var. *hortensis*—belonged to different clusters of the phylogenetic tree. This could be ascribed to the frequent hybridization between cultivated and wild radish species, which was also suggested based on chloroplast SSR data and mitochondrial DNA types (Yamagishi and Terachi 2003; Yamagishi and Sasaki 2004). The European wild species are distributed in all three clusters, so that the authors could not conclude which wild species is the ancestor of the cultivated radish (Lu et al. 2008). Allozyme variation was studied within and among open-pollinated cultivars of radish and wild radish populations and it was concluded that domesticated radish retained a population structure similar to that of wild populations, which again suggests frequent hybridizations between wild and cultivated radish (Ellstrand and Marshall 1985). Yamagishi and Sasaki (2004) analyzed 36 radish accessions for variation in the mitochondrial 3′region of *ATP6*, and found that 75% of the varieties were assigned to type I or type II, while seven varieties contained both type I and type II and further varieties with type III or IV were rare. East Asian wild radish was related to European wild radishes, which made the authors suggest that the East Asian wild radish originated in Europe as a weed, as suggested by many others. Kim et al. (2007) identified a novel mitochondrial genome type and developed molecular markers to classify radish cytoplasm types. In addition to the *orf-138-atp8* Ogura mitochondrial DNA organization of cytoplasmic male sterility type, they detected several sub-stoichiometric organizations linked to the *atp8* gene. Using the classification system based on PCR amplification, the authors could assign a total of 90 radish cultivars or accessions into three different mitotypes. Lee et al. (2009) recently investigated the organization *atp6-nad3-rps12* of the mitochondrial DNA of a novel cytoplasmic male sterility (CMS) type (DCGMS). They identified an SNP that lead to a frame-shift mutation, which was diagnostic for the DCGMS type.

Several reports address the genetic variation of wild, so-called weedy radish (*R. raphanus*) and *R. raphanistrum*. Panetsos and Baker (1968) analyzed several populations of *Raphanus* growing in California for both morphological and cytological traits and concluded that the populations of the highly successful “wild” (weedy) *R. sativus* originated by hybridization
Genetics, Genomics and Breeding of Vegetable Brassicas

between the cultivated forms of radish with the introduced weed species *R. raphanistrum*. Campbell and Snow (2009) concluded that feral radish likely evolved through gene-flow from *R. raphanistrum* to cultivated radish based on experiments with radish populations over three generations. Based on chloroplast DNA analyses (*trnL-rpl32*), bidirectional hybridization between the progenitor species was suggested for the creation of wild/feral radish in California (Ridley et al. 2008). By studying plant growth traits, it was suggested that loss of the rosette and early flowering were key adaptations in the success of wild radish as a major agricultural weed (Sahli et al. 2008). Huh and Ohnishi (2002) assessed the genetic diversity and genetic structure of East Asian wild radish populations (*Raphanus sativus* var. *hortensis* f. *raphanistroides*), a natural and a cultivated population of *Raphanus sativus* var. *hortensis* from Kazakhstan and a *R. raphanistrum* population from Russia using AFLPs. They showed that the variation within populations was very large (97%) and that there was a close genetic relationship between *R. sativus* from Kazakhstan and *R. raphanistrum*, which confirmed the assumption that *R. raphanistrum* played a role in the origin of *R. sativus*, which probably took place somewhere in the eastern Mediterranean as described previously. In another study, the same authors (Huh and Ohnishi 2001) studied the diversity and population structure of Japanese and Korean populations of wild radish (*R. sativus* var. *hortensis* f. *raphanistroides*) that mainly grow on beaches using allozyme diversity. The genetic diversity within populations was high, which they explained by the insect pollinated outcrossing breeding system, gene-flow from cultivated radishes, large population sizes and high fecundity. In a dendrogram, the Japanese and Korean clusters were clearly distinguished with a close relationship between geographical origin and genetic distance, also within the Korean and Japanese clusters.

3.7 Use of Genetic Diversity Assessments for Germplasm Conservation and for Practical Breeding

The conservation of plant genetic resources in genebanks involves the selection of accessions to be conserved and the maintenance of these accessions for current and future users. For both these issues, knowledge is needed about the distribution of genetic diversity within and between accessions to be sampled from the gene pool, but also about the changes in variation of these samples as a result of regenerations (van Hintum et al. 2007). The goal of genebanks is to preserve the levels of variability in traits such as shape, color, taste, pest and disease resistances and seasonal growing type among other properties that can be valuable sources for breeding. Evaluation based solely on morphological characteristics can
Diversity Analysis and Molecular Taxonomy of Brassica Vegetable Crops

be insufficient for an accurate characterization of accessions, since it may under- or over-estimate genetic diversity between accessions. This is clearly illustrated by the fact that especially in Brassicas, convergent evolution in almost every trait studied hampers classifications. This is obvious in the studies that profile molecular markers for different vegetable mustards (Brassica juncea), since the different vegetable types clustered together and similar types were distributed over different clusters (Ren et al. 1995; Qiao et al. 1998; Fu et al. 2006; Qi et al. 2007, 2008). In Brassica rapa, turnips from the East and West also formed distinct clusters when grouped according to AFLP profiles (Zhao et al. 2005). Molecular marker profiles can assist in variety identification (Hu and Quiros 1991), genebank management like elimination of duplicated entries and checking for uniformity and stability within accession (Kresovich et al. 1992; Farnham 1996; Phippen et al. 1997; van Hintum et al. 2007), taxonomic studies (Demeke et al. 1992; Margale et al. 1995; Lanner-Herrera et al. 1996) and development of minimal sets of markers like RAPD primers, ISSR or SRAP primers to be able to distinguish all cultivars (Srivastava et al. 2001; Tonguç and Griffiths 2004; Louarn et al. 2007; Wang et al. 2008). The latter can be used in law suits, where claims on origin of accessions/cultivars can be tested. All these studies corroborate to the usefulness of molecular marker methods to evaluate changes over regeneration processes and the genetic differentiation among accessions.

Genetic diversity assessments can also be used to improve breeding programs. Because Brassica vegetables and also oilseeds have been cultivated for many centuries in different parts of the world, the variation within the species has increased as a result of ongoing breeding. Different Brassica crop breeding programs have different objectives and priorities since each vegetable type is characterized by its own characteristics and the growing areas diverge. The relationship between accessions revealed by diversity studies, as mentioned above, indicated that different cultivar groups are often region specific, and have emerged independently in the different geographical regions. The market demands are considered by breeders in designing the most desirable ideotype, like Chinese cabbage with ovate or cylindrical heads is favored in different geographical regions and within an ecological region, breeding for specific types (like the ?pak choi related rosette shaped wutacai and early flowering caixin) is necessary to meet specific uses.

Molecular analysis has shown that there is considerable genotypic variation in the B. rapa gene pool, while several studies on B. juncea mustard oilseeds and R. sativus in Pakistan, but also on canola quality B. juncea in Canada and Australia, conclude that genetic variation is limited (Rabbani et al. 1998a, b, c; Burton et al. 2004; Khan 2008). The genetic diversity studies, using both molecular markers and phenotypic descriptors, can assist breeders in the choice of starting material for breeding programs, in
broadening the base of genetic variation in their gene pools, and in selecting parents for breeding programs or in generating heterotic pools for hybrid breeding programs.

Acknowledgements

We thank Dr. Agnieszka Ludwikow for reviewing the chapter and Dr. Ningwen Zhang for critically reading the manuscript. This work was co-financed by the Centre for BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative/Netherlands Organisation for Scientific Research. Dunia Pino Del Carpio was funded by an IOF genomics program “Brassica nutrigenomics” and Jianjun Zhao by the programme Strategic Scientific Alliances (PSA) of the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Chinese Ministry of Science and Technology.

References


Diversity Analysis and Molecular Taxonomy of Brassica Vegetable Crops


De Candolle A (1886) Origin of Cultivated Plants. 2nd edn. Hafner, New York, USA.


Lanner C (1998) Relationships of wild *Brassica* species with chromosome number 2n=18, based on comparison of the DNA sequence of the chloroplast intergenic region between trnL (UAA) and (GAA). *Can J Bot* 76: 228–237.


Diversity Analysis and Molecular Taxonomy of Brassica Vegetable Crops


Diversity Analysis and Molecular Taxonomy of Brassica Vegetable Crops


Molecular Linkage Maps: Strategies, Resources and Achievements

Danuta Babula-Skowrońska,1* Agata Cieśla,1 and Jan Sadowski1,2

ABSTRACT

This chapter provides a detailed overview of marker systems, genetic maps and the status of genome mapping for diploid species of the Brassica genus. Progress in molecular biology techniques development and a high level of DNA polymorphisms, disclosed both among various Brassica accessions and subspecies, allowed construction of comprehensive genetic maps. We describe different marker systems used in the Brassica chromosomal maps construction with their specific advantages and disadvantages, thus the choice of a particular marker system must be adapted to the particular problem to be solved. On the other hand, the presence of the same markers on different Brassica maps allows comparison of their localization, giving insight into the structure and evolution of genomes of the diploid and amphidiploid species. Construction of Brassica genetic maps appeared to be of great significance for increasing breeding efficiency as it permitted detection and analysis of QTLs, marker- or gene-assisted selection, gene flow and fingerprinting for clone identification of economically important genes. Furthermore, the maps facilitated detection of a target region by searching public mapping database. Importantly, the Brassica species are closely related to Arabidopsis thaliana, therefore the knowledge obtained from the A. thaliana genome sequencing has opened the way for comparative genomics, evolutionary studies and gene structure analysis. Finally, it needs to be emphasized that all genetic mapping experiments are based on polymorphism of genetic markers, thus the final pattern

1Institute of Plant Genetics, Polish Academy of Sciences, Strzeszyńska 34, 60-479, Poznań, Poland.
2Department of Biotechnology, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland.
*Corresponding author: dbab@igr.poznan.pl
of duplication/triplication of the brassica genomes is incomplete and requires further verification. Nevertheless, the present knowledge of genomic relationship among diploid brassica species, collected due to genetic maps development, made the understanding of the brassica genomes evolution possible. Finally, the studies of chromosomal location and expression of duplicated genes in the brassica species will provide insight into the role of polyploidization/diploidization cycles in divergence of gene functions leading to the phenotypic plasticity of vegetable brassica species.

**Keywords:** DNA markers, linkage maps, comparative maps, segmental duplications, paleopolyploids

### 4.1 Introduction

The first genetic linkage associations of genes controlling morphological traits in brassica were detected over 80 years ago (Pease 1926). Further research describing linkage groups for the brassica species concerned only a limited number of morphological and isozyme markers (Yarnell 1956; Stringam 1977; Sampson 1978; Arus and Orton 1983). Despite identification and rough mapping of genes for morphological characters and isozyme markers, little was known about their organization in the brassica genomes (James and Williams 1980). Development of molecular biology techniques and a high level of DNA polymorphisms disclosed both between various brassica accessions and subspecies allowed construction of comprehensive brassica genetic maps (Figdore et al. 1988; Song et al. 1988; McGrath et al. 1990). The first detailed restriction fragment length polymorphism (RFLP) maps of brassica have been published in the beginning of the 1990s (McGrath et al. 1990; Song et al. 1991; for a review see Babula et al. 2007). Many maps of brassica have been constructed on the basis of the F2 or doubled haploid (DH) lines derived from different species, subspecies or varieties. This facilitated mapping of the genes or quantitative trait loci (QTLs) of economically important traits, identification of introgressions between different genomes, marker-assisted breeding and gene pyramiding aimed at crop improvement. The presence of the same markers on different brassica maps allows comparison of their localization, giving insight into the structure and evolution of genomes of the diploid and amphidiploid species.

Up to date more than 1,000 different markers have been located on 15 genetic maps of the brassica genus species using different populations and marker systems (Lim et al. 2007). This chapter provides a detailed overview of genetic maps and the status of genome mapping for diploid species of brassica genus.
4.2 Evolution of Marker Types: From RFLPs to SNPs

Along with accumulating knowledge about primary genome structure, new types of markers, namely molecular markers, have appeared. Since the polymerase chain reaction (PCR) technique was first introduced in 1983 by Kary Mullis, a range of new PCR-based markers have been used in genetic map construction. The DNA markers have shed new light on plant genome research and breeding. The molecular markers have been widely used for identification and analysis of agronomically important genes (Plieske and Struss 2001) and for genetic analysis of the *Brassica* genomes structure (Williams et al. 1990; Table 4-1). Importantly, the *Brassica* species are closely related to *Arabidopsis thaliana*, therefore the knowledge obtained from the *A. thaliana* genome sequencing has opened the way for comparative genomics, evolutionary studies and gene structure analysis (Snowdon and Friedt 2004). As a result of the genome sequencing of the model plant *A. thaliana* (The Arabidopis Genome Initiative 2000), many techniques of marker production used in the *Brassica* research have become more informative. Each marker system has specific advantages and disadvantages, thus the choice of a particular marker system must be adapted to the particular problem to be solved. Molecular markers widely used in *Brassica* research can be classified into three groups as listed below.

1. Hybridization-based DNA markers, such as restriction fragment length polymorphism (RFLP)
2. PCR-based DNA markers, such as random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), sequence-characterized amplified region (SCAR), cleaved amplified polymorphic sequences (CAPS), simple sequence repeat (SSR) and sequence tagged site (STS), and sequence-related amplified polymorphism (SRAP)
3. DNA markers based on single nucleotide polymorphism (SNP; for details about molecular markers see Gupta et al. 1999).

4.2.1 Restriction Fragment Length Polymorphism

Restriction fragment length polymorphism (RFLP) is one of the earliest DNA-based techniques used in genetic studies. The polymorphism is analyzed in Southern blot hybridization assay. It was first used for human genome mapping and later it was adopted for plant genome mapping, including *Brassica* (Table 4-1; for a review see Gupta et al. 1999). There were two main types of probes used in the RFLP technique. In the first, unknown genomic sequences were used as hybridization probes; they were fragments of genomic DNA obtained after digestion with restriction endonucleases. The enzyme used most often was *PstI* and the restriction fragments obtained were referred to as *PstI*-fragments. With progress in
Table 4-1 Characteristics of genetic markers applied in the *Brassica* studies.

<table>
<thead>
<tr>
<th>Properties</th>
<th>RFLP</th>
<th>RAPD</th>
<th>AFLP</th>
<th>SCAR</th>
<th>CAPS</th>
<th>SSR</th>
<th>STS</th>
<th>SRAP</th>
<th>SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of polymorphism</td>
<td>point mutations, insertions, deletions, high</td>
<td>point mutations, insertions, deletions, high</td>
<td>point mutations, insertions, deletions, high</td>
<td>point mutations, insertions, deletions, high</td>
<td>point mutations, insertions, deletions, high</td>
<td>changes in microsatellites sequence length mean</td>
<td>point mutations, insertions, deletions rare</td>
<td>point mutations, insertions, deletions high</td>
<td>point mutations high</td>
</tr>
<tr>
<td>Frequency of occurrence in genome</td>
<td>high</td>
<td>high</td>
<td>high</td>
<td>high</td>
<td>high</td>
<td>high</td>
<td>mean</td>
<td>high</td>
<td>very high</td>
</tr>
<tr>
<td>Genome location</td>
<td>coding regions</td>
<td>whole genome</td>
<td>whole genome</td>
<td>whole genome</td>
<td>whole genome</td>
<td>repeated DNA</td>
<td>whole genome</td>
<td>coding regions</td>
<td>whole genome</td>
</tr>
<tr>
<td>Type of inheritance</td>
<td>codominant</td>
<td>dominant</td>
<td>dominant</td>
<td>dominant</td>
<td>codominant</td>
<td>codominant</td>
<td>dominant</td>
<td>dominant/ codominant</td>
<td>domoninant/ codominant</td>
</tr>
<tr>
<td>Possibility of alleles detection</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes/yes (rare)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>DNA sequence knowledge required</td>
<td>no</td>
<td>no</td>
<td>yes/no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
Molecular Linkage Maps: Strategies, Resources and Achievements

129

Genome sequencing, the expressed sequence tags (ESTs) could be used as hybridization probes in RFLP analysis.

RFLP analysis was widely used in the *Brassica* genetics. However, taxonomists were the first beneficiaries of RFLP. Random nuclear sequences selected from *PstI* genomic DNA library were used to estimate the phylogenetic relationships between vegetable subspecies of *Brassica oleracea* and *Brassica rapa* (Song et al. 1990). The RFLP markers have been found to be very useful in genetic map construction of the *Brassica* genomes due to the high level of DNA polymorphism in these species, and the codominant inheritance of RFLP markers. Since the 1980s many RFLP-generated maps have been constructed for different *Brassica* crops. The first genome restriction fragment polymorphisms for the *Brassica* species were estimated by screening *EcoRI*, *HinIII*, or *EcoRV*-digested total genomic DNA (Figdore et al. 1988). The first linkage map based on RFLP for *B. oleracea* was constructed by Slocum et al. (1990). These studies proved that with the help of RFLP markers, it was possible to detect genome rearrangements such as tandem duplications and a strong conservation of some parts of the linkage loci organization. RFLP markers were used in linkage map construction for different vegetable *Brassica* as shown by Landry et al. (1992), Camargo et al. (1997) or Hu et al. (1998). RFLP markers were used also in integration of linkage maps based on intraspecific and interspecific mapping populations of *B. oleracea* (Kianian and Quiros 1992a). It significantly increased the level of polymorphism and introduced RFLP analysis as an efficient method for evolutionary considerations. The RFLP together with AFLP markers were used in map construction of the *B. oleracea* and *B. juncea* chromosomes (Sebastian et al. 2000; Pradhan et al. 2003). This combined approach offered the advantages of both marker types.

The sequencing project of the *A. thaliana* genome has opened a new era in the *Brassica* genetics. In the absence of complete knowledge about the *Brassica* genome, ESTs and genome sequence of related organisms have offered an opportunity to examine a large number of coding sequences (Table 4-2). EST library from the *B. rapa* ssp. *pekinensis* flower buds was used for characterization of the coding-sequence divergence of closely related *B. rapa* and *A. thaliana* genomes (Tiffin and Hahn 2002). These studies showed that the sequences, which were not significantly similar to those of *A. thaliana*, did not demonstrate essential similarity to other sequences deposited in public databases. This observation suggests that some of the similar sequences that were not detected for being significantly changed was a result of distinct selection in comparison to those of *A. thaliana*. It shows that *A. thaliana* ESTs can be used in evolutionary studies of *Brassica*, but the most popular application of this type of hybridization probes is map construction. ESTs refer to coding sequences so they can be used for identification of coding regions in a genome and for discovering
new genes. Thus, their localization on the genetic maps can bring valuable information. RFLP markers based on unknown sequences as well as ESTs were used in the construction of the comparative map between *B. oleracea* and *A. thaliana* (Lan et al. 2000). Andy Paterson’s group has used ESTs from *A. thaliana* in *Brassica* mapping and this approach allowed determination of the structural changes in the genome of these two species (Lan et al. 2000). ESTs can be also very useful for localization of the disease resistance loci on the genetic maps in economically important *Brassica* species, and regarding the conservation of ESTs, the same probes can be used in genome studies even in distantly related species like *Arabidopsis*, *Gossypium* and *Sorghum* (Paterson et al. 1996; Lan et al. 2000).

ESTs from the *A. thaliana* genome sequencing project served as probes for RFLP analysis in map construction for *B. oleracea*. EST-based probes corresponded to 12 main functional categories of known genes from *A. thaliana* (Babula et al. 2003). A large number of RFLP markers were used for construction of the sequence-tagged linkage map for *B. rapa*. ESTs clones from four different tissue-specific libraries of *B. rapa* and sequences representing genes from *A. thaliana* generated by PCR, were used as hybridization probes (Kim et al. 2006). All markers used in this experiment made the resultant map very detailed.

### 4.2.2 Random Amplified Polymorphic DNA

Interest in the development of genetic maps has prompted the search for markers that would allow fast and cheap genome analysis. One of the markers proposed was the random amplified polymorphic DNA (RAPD; Williams et al. 1990). This is a simple marker system based on the amplification of genomic DNA with single oligomer primers of random sequence. The primers permit detection of polymorphisms when DNA sequence is not known. It is possible to analyze a wide range of species using a universal set of primers (for more details about RAPD markers see Williams et al. 1990).

---

**Table 4-2** Number of the *Brassica* ESTs clones available in EST database in NCBI (updated May 2009).

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of ESTs</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Brassica rapa</em></td>
<td>194,741</td>
</tr>
<tr>
<td><em>Brassica nigra</em></td>
<td>1,814</td>
</tr>
<tr>
<td><em>Brassica oleracea</em></td>
<td>63,201</td>
</tr>
<tr>
<td><em>Brassica carinata</em></td>
<td>2,500</td>
</tr>
<tr>
<td><em>Brassica juncea</em></td>
<td>5,393</td>
</tr>
<tr>
<td><em>Brassica napus</em></td>
<td>656,151</td>
</tr>
</tbody>
</table>
RAPD markers have become very useful in the *Brassica* species research taking into account the high level of polymorphism within each species, and the various genomes in basic diploid and natural amphidiploid species. *Brassica* genomes’ analysis revealed that with the use of a blend of arbitrary primers it is possible to identify genome-specific and species-specific RAPD markers (Quiros et al. 1991; Table 4-1). RAPD markers proved to be important in investigation of genetic identities and relationships among botanical varieties of *B. oleracea* (Kresovich et al. 1992). RAPD together with RFLP markers were used to analyze the relationships between selected Brassicaceae species. A range of 18 different botanical varietes of *Brassica* species and one variety of *Raphanus sativus* was chosen as material for taxonomic analysis. Results revealed the worth of RAPD markers for this type of analysis (Thormann et al. 1994). Lazaro and Aguinalde (1998) used RAPD markers to evaluate genetic diversity in *B. oleracea* and wild relatives. As a result of these studies, the *Brassica* populations were grouped into three main taxonomic branches. RAPDs were used for *B. rapa* linkage map construction and for selected QTLs analysis. This investigation revealed that some of the RAPD markers used were linked with agronomic traits such as bolting time and self-incompatibility, which makes RAPD markers useful in marker-assisted selection (Nozaki et al. 1997).

### 4.2.3 Amplified Fragment Length Polymorphism

Amplified fragment length polymorphism (AFLP) is a technique resembling that of RFLP. The main difference between these two approaches is that AFLP is based on PCR amplification instead of southern hybridization for detection of restriction fragments, like in RFLP technique (for details about AFLP markers see Vos et al. 1995; Mueller and Wolfenbarger 1999).

AFLP markers have been used in many different studies, e.g. phylogenetic relationships, systematic studies or epidemiological hospital monitoring in a wide range of taxa (Muller and Wolfenbarger 1999). The AFLP analysis has become a tool successfully adapted in the *Brassica* genomics. Because of its many advantages, application of AFLP markers permits fast construction of genetic linkage maps in different members of the *Brassica* genus like in the case of *B. juncea* (Pradhan et al. 2003), *B. oleracea* (Sebastian et al. 2000) or *B. napus* (Lombard and Delourme 2001) (Table 4-1).

Literature data show some interesting features of AFLP markers. In construction of the *B. juncea* linkage map, differences between *Eco*RI- and *Pst*I-based AFLP distributions were observed (Pradhan et al. 2003). *Eco*RI-based AFLP analysis was more disposed to form some clusters than *Pst*I-based AFLP markers (this was also noted in tomato and maize mapping). In connection with these observations, it is presumed that *Eco*RI-based AFLPs
are more suitable for map saturation, while *Pst*I-based AFLP markers have a better genome distribution. AFLP markers were also used for integration of the genetic maps across different populations in *B. oleracea* (Sebastian et al. 2000). They were useful in filling gaps in the integrated map, but it was noticed that a significant number of AFLP markers used was located near the centromer regions. This finding can be explained by the repetitive DNA presence in these regions (Sebastian et al. 2000).

The main disadvantage of AFLP markers is that they mostly represent the dominant type of inheritance. Also methylation of genomic DNA can result in false positive polymorphism when the restriction enzymes are sensitive to methylation (Li and Quiros 2001). AFLP technique has been widely used to target specific loci but the assay is too complicated to be adapted for high-throughput selection, marker-assisted selection (MAS) or map-based cloning. This particular marker, AFLP, can be converted into a PCR-based one, such as CAPS or SCAR (Piao et al. 2004).

4.2.4 Sequence-Characterized Amplified Region and Cleaved Amplified Polymorphic Sequence

These two types of PCR-based markers were especially useful in the *Brassica* genome analysis when special agriculturally important genes were searched (Table 4-1). Clubroot disease caused by *Plasmodiophora brassicae* is a very dangerous infection of the *Brassica* crops including Chinese cabbage (*B. rapa* ssp. *pekinensis*) (Piao et al. 2004) (for more information about clubroot disease in the *Brassica* crops see Chapters 2 and 5). The source of resistance in *Brassica* has been identified as *CRb* gene; it has been mapped using AFLP markers. A few AFLP markers, which mapped near the resistance gene, were selected for conversion into sequence-characterized amplified region (SCAR) and cleaved amplified polymorphic sequence (CAPS) markers. AFLP markers were sequenced from both ends and specific primers served for development of SCAR markers. SCAR markers did not give polymorphic products, so the products obtained were digested by restriction enzymes. With the help of this kind of markers a local map around *CRb* gene was constructed.

*B. juncea* known as Indian mustard is a very important crop in China, however, it is often affected by white rust, a dangerous crucifers disease caused by *Albugo candida*. Breeding for resistance against this pathogen has not been successful although sources of resistance are available. The gene responsible for white rust resistance was mapped for *B. rapa* (Kole et al. 1996, 2002) and *B. juncea* (Prabhu et al. 1998, Mukherjee et al. 2001; Varshney et al. 2004), but RFLP and RAPD markers used for the target gene are not appropriate for marker-assisted selection (MAS). Therefore, it was necessary to develop markers more tightly linked with pathogen resistance gene and
produce SCAR/CAPS markers (Varshney et al. 2004). CAPS markers were generated from RAPD markers by the way described above.

SCAR and CAPS markers are very useful in map-based cloning (MBC) and marker-assisted selection (MAS). The markers of these types can be also used for screening plants for specific genes in breeding programs. SCAR and CAPS markers are quite fast in the search for resistance genes and importantly, many markers of this type are codominant.

### 4.2.5 Simple Sequence Repeats

The simple sequence repeats (SSRs), known as microsatellites, are short DNA sequences in which a motif of one to six bases is tandemly repeated. Their origin is closely related to the errors appearing during DNA replication and recombination process (Goldstein and Schlötterer 1999). Subsequently, additional sets of these repeated sequences are added to the strand. Many studies indicate that SSRs are abundant and their distributions differ not only among species but also between transcribed and non-transcribed regions both in plants and mammals (Katti et al. 2001; Morgante et al. 2002; Toth et al. 2000). The analyses of plant DNA sequences revealed a frequency of one SSR every 6–7 kb, which is equivalent to the corresponding data for mammals, while the studies based on oligonucleotide hybridization suggested the presence of one SSR every 65 to 80 kb (Panaud et al. 1996; Cardle et al. 2000). On the other hand, analysis of the microsatellite motifs distribution in the *B. oleracea* genomic shotgun sequences brought the frequency of one SSR every 4 kb (Iniguez-Luy et al. 2008).

SSRs have many characteristic features such as: a high level of polymorphism resulting from length variability, codominant Mendelian inheritance, almost random distribution in eukaryotic genomes, locus-specificity, conservation among closely related species and relative easiness of isolation and characterization (Weber and May 1989; Beckmann and Soller 1990; Queller et al. 1993). Hence, they appeared as a basic marker type for construction of the genetic maps (for example Dietrich et al. 1994; Dib et al. 1996; McCouch et al. 2002; Piquemal et al. 2005; Suwabe et al. 2006). Furthermore, SSR markers were used in linkage and association analyses (Abdurakhmonov et al. 2005), phylogenetics (Plieske and Struss 2001; Chung and Staub 2003; Flannery et al. 2006) and population studies (Rosenberg et al. 2002). Because of the duplicated nature of many genomes, the microsatellite sequences could occur in multiple copies with similar flanking regions forming families (Zhang 2004). Thus, many primer pairs could amplify two or more loci. On the other hand, identification of the microsatellite motifs within gene sequences made it possible to distinguish their duplicated copies (Babula-Skowronska et al., unpublished).
This wide range of SSR use in genetic studies initiated development of microsatellite discovery programs for the majority of agronomically important crops (Condit and Hubbell 1991; Akkaya et al. 1992). Currently, in the public domain 1,545 SSRs for *Brassica* are available, which have been isolated from different *Brassica* genomes or transferred from *A. thaliana* (Szewc-McFadden et al. 1996; Westman and Kresovich 1998; Plieske and Struss 2001; Saal et al. 2001; Lowe et al. 2002; Suwabe et al. 2002; Table 4-3). Identification of SSRs in the *Brassica* genomes is based on probe hybridization to genomic or cDNA libraries followed by DNA sequencing (Kresovich et al. 1995; Plieske and Struss 2001; Suwabe et al. 2002; Lowe et al. 2004). Recently, SSRs were identified for *B. rapa* based on the bacterial artificial chromosome (BAC)-end sequences selected from the KBrH library (Choi et al. 2007). A dynamic increase in the number of SSR sequences in the *Brassica* database allowed development of many genetic markers, which were added to the framework RFLP maps or became the basis of new SSR-based linkage maps (Sebastian et al. 2000; Piquemal et al. 2005; Choi et al. 2007; Gao et al. 2007). As SSR markers are highly transferable with high reliability among related species, they are used for integration of genetic maps. The first experiments concerned integration data from two different segregating populations of *B. oleracea* (Sebastian et al. 2000; Babula-Skowronska et al., unpublished) and integration of A-genome maps in *B. napus* and *B. rapa* (Choi et al. 2007). Moreover, SSRs developed from the *B. oleracea* sequences were used for inspection of genetic relationships within the clade (Tonguc and Griffiths 2004) and in chloroplast genomes of this species and wild relatives (Allender et al. 2007), while the SSRs from *B. napus* were used for finding their distribution in Brassicaceae such as *B. oleracea*, *B. rapa*, *B. nigra*, Diplolotaxis ssp., *B. tournefortii*, *Sinapis alba*, *Raphanus sativus* and *Eruca sativa* and for estimation of the extent of genetic diversity among 32 varieties of *B. napus* (Plieske and Struss 2001). These data indicated high variation potential of the species of U triangle, but low transferability of microsatellite markers into distantly related Brassicaceae species. Also, the distribution, putative function, and mutational mechanisms of SSRs’ evolution have been studied (Ayele et al. 2005; Li et al. 2002; Iniguez-Luy et al. 2008).

### 4.2.6 Sequence Tagged Sites

Sequence tagged sites (STS) belong to PCR-based markers, characterized by a high specificity. There are supposed to be only a single locus in a genome for particular STS, if properly selected. Markers of this type are often used to increase density of genetic maps. The reference linkage map for the *Brassica* A genome developed in the multinational *B. rapa* sequencing project, also includes some STS markers (Choi et al. 2007; Table 4-1). The primer pairs for
Table 4-3 Summary of the *Brassica* SSR markers based on the published information and data available in the public domain (www.brassica.info).

<table>
<thead>
<tr>
<th>Source of of SSR data</th>
<th>References</th>
<th><em>B. rapa</em></th>
<th><em>B. nigra</em></th>
<th><em>B. oleracea</em></th>
<th><em>B. napus</em></th>
<th><em>A. thaliana</em></th>
<th>Total number of SSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBSRC Microsatellite</td>
<td>Lowe et al. 2002; Lowe et al. 2004</td>
<td>89</td>
<td>113</td>
<td>95</td>
<td>101</td>
<td></td>
<td>397</td>
</tr>
<tr>
<td>Program</td>
<td>HRI Microsatellites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Smith and King 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Kresovich et al. 1995; Szewc-McFadden et al. 1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Lagercrantz et al. 1993</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>AAFC BAG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>Plieske and Struss 2001; Saal et al. 2001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Suwabe et al. 2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Uzanova and Ecke 1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Bell and Ecker 1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>INRA Versailles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Celera AgGen <em>Brassica</em> Consortium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>372</td>
</tr>
<tr>
<td></td>
<td>Piquemal et al. 2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FITO (Wisconsin)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>587</td>
</tr>
<tr>
<td></td>
<td>Choi et al. 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>470</td>
</tr>
</tbody>
</table>

Molecular Linkage Maps: Strategies, Resources and Achievements 135
STS markers were designed from the mapped genomic sequences (Teutonico and Osborn 1994) and BAC-end sequences.

To overcome limitations connected with RFLP and AFLP markers, such as high costs and time consuming procedures or not enough markers associated for genes crucial in breeding programs, a way should be found to transform RFLPs and AFLPs into PCR-based markers, such as STS. This was found in studies on finding markers associated with *Phoma* resistance genes in the *Brassica* B genome (Plieske and Strus 2001).

### 4.2.7 Sequence-Related Amplified Polymorphism

Sequence-related amplified polymorphism (SRAP) is a simple PCR-based marker system developed by Li and Quiros (2001) (Table 4-1). It is based on amplification of genomic sequences with two specific primers (for details see Li and Quiros 2001). SRAP marker technique allows for coding sequence investigation and easy sequencing of selected bands. Markers of this type have a good distribution in the genome because they are based on open-reading frames (ORFs), so they have been used in the *B. oleracea* map construction together with AFLP markers (Li and Quiros 2001). SRAP markers were placed in the nine main linkage groups of *B. oleracea* resulting in no significant difference in genome coverage obtained for AFLP. SRAP markers revealed relatively high frequency of codominant type of inheritance, which is a very important attribute. Using the resultant map, SRAP markers were used for tagging of glucosinase desaturation gene (*BoGLS-ALK*). SRAP133 marker was located inside this gene that is important for MAS. A MAS allows fast screening of numerous segregating populations. SRAP markers have also many other applications. They can generate high level of polymorphisms in many other crops and they were used to find markers for male sterility in Chinese cabbage, *cms* fertility restorer gene in rapeseed and virus resistance gene in celery (Li and Quiros 2001). SRAP markers have also found application in cDNA fingerprinting. In *B. rapa* some tissue-specific bands were identified with markers of this type, which corresponded to genes expressed only in the tissue investigated (Li and Quiros 2001).

SRAP markers were also used in investigation of markers for seed coat color gene in *B. rapa*. The color of the *Brassica* seeds varies from dark brown to bright yellow. The yellow color of the seeds is most desirable as the seed coat in this color is thinner and yellow seeds contain more oils, more proteins with lower fiber content, which is important for its use as fodder. Thus, it would be important to identify genes responsible for seed coat color. Three such genes were identified in *B. rapa*. SRAP marker linked with the seed coat color gene (*Br1/br1*) was found. On the basis of this SRAP marker, the SNP and SCAR markers were obtained. The latter can be
useful in MAS for monitoring genotypes in the early stage of development and in selection of yellow seeds coat color alleles configuration expected (Rahman et al. 2007).

4.2.8 Single Nucleotide Polymorphism

Single nucleotide polymorphisms (SNPs) resulting from point mutations in DNA sequence occur abundantly in the genome. Analysis of SNPs allows identification of allelic variation of particular genes. SNPs are widely used in human disease diagnostics, they have a great potential in genetic mapping and genome analysis in many different organisms also in vegetables Brassica (Snowdon and Friedt 2004; Table 4-1).

SNP markers were used in the comparative genomics study of B. rapa and A. thaliana to screen a specific region responsible for clubroot resistance in Brassica (Suwabe et al. 2006). SNPs were also used to analyze the seed color genes in F₂ and BC₁ of B. rapa breeding materials (Rahman et al. 2007).

In future, SNPs are expected to be an important marker system in the Brassica genetic studies. Knowledge of genome sequences of Brassica could be used to develop locus-specific SNP markers. Before it happens, it is necessary to develop tools and methods for SNP detection and interpretation in polyploid genomes of Brassica. New high-throughput methods like matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) may be adapted to SNPs detection. It allows reduction in the time and costs of SNP analysis (Snowdon and Friedt 2004).

4.3 Development of Public Marker Resources for Brassica

Together with the start of the Multinational Brassica Genome Project (MBGP Brassica) the information platform at the web pages has been made accessible (www.brassica.info, http://brassica.bbsrc.ac.uk). The platform facilitates exchange of information from the Brassica genome analyses at the genetics, genomics, proteomics and metabolomics levels between different laboratories. Besides the above data, the platform also includes the sets of genetic markers mapped in the hitherto published Brassica linkage maps. Currently, a total of 2,665 Brassica sequence tagged markers including 584 RFLPs, 1,545 SSRs, 536 SNP/InDel, intron polymorphism and ACGM plus over 1,600 additional AFLP and RAPD markers have been accumulated. Additionally, the detailed description for microsatellite sets has been placed at the microsatellite information exchange http://www.brassica.info/resource/markers/ssr-exchange.php. Here, the basic information concerning primer sequences, repeat classes and resources has been given. The majority of SSRs have been isolated and described by BBSRC Microsatellite Program,
Celera Consortium (commercial), INRA Versailles and AAFC Consortium (commercial). This characterizes predominantly SSRs from *B. rapa*, *B. nigra*, *B. oleracea* and *B. napus*. The remaining SSRs have been developed and characterized by several small research groups (Lagercrantz et al. 1993; Bell and Ecker 1994; Kresovich et al. 1995; Szewc-McFadden et al. 1996; Uzanova and Ecker 1999; Saal et al. 2001; Plieske and Struss 2001). Many of the isolated SSRs have been localized on the *Brassica* genetic maps (for example Sebastian et al. 2000; Piquemal et al. 2005; Choi et al. 2007; Gao et al. 2007; D. Babula-Skowronska et al., unpubl). Increased mapping data covering markers for the existing *Brassica* linkage maps, dramatically enriched this database.

4.4 Nomenclature of Genetic Markers and Genes

The marker names, accessions and orientation of the linkage groups in the *Brassica* species were arbitrary, because different nomenclatures for markers, genes and their paralogs were used by different researchers (Østergaard and King 2008). Hence, many markers with different names located on the different linkage maps might represent the same locus. The lack of standardized nomenclature for linkage groups and common probes in the *Brassica* species can hinder the exchange of genetic information and comparison of the previously published maps. Therefore, the Steering Committee for the Multinational *Brassica* Genome Project recommended a new nomenclature system for the *Brassica* linkage groups, markers and genes ([http://brassica.info/resource/maps/lg-assignments.php](http://brassica.info/resource/maps/lg-assignments.php); Østergaard and King 2008). The “single letter marker type designator codes” system for marker names has been adopted after De Vicente et al. (2004). According to this proposition, the single letter marker assay type designate: a-AFLP, e-EST, f-RFLP, m-SSR, p-CAPS, r-RAPD and t-STS, c-SCAR, s-SNP, x-other and u-unspecified. Østergaard and King (2008) proposed a system of naming the linkage groups assigned to the three diploid *Brassica* genomes. Here, the consensus nomenclature of the chromosome/linkage group (A1–A10, C1–C9 and B1–B8 in *B. rapa*, *B. oleracea* and *B. nigra*, respectively) has been assigned to the canonical diploid *Brassica* genomes in the “triangle of U” as: A, B and C. It should facilitate integration of genetic linkage maps constructed by different research groups using different mapping populations and different sets of molecular markers. The standard systematic gene nomenclature system for the *Brassica* genus takes the form: `<GENUS 1 LETTER> [<species 2 letters>]<GENOME 1 LETTER> | <X>.<NAME 3–6 LETTER CODE>,<locus assignment 1 letter>`. An example is the orthologue of the *Arabidopsis* ABA INSENSITIVE 1 (*ABI1*) gene isolated from the C genome of *B. napus*: *BnaC.ABI1.a*. 
4.5 Mapping Populations

4.5.1 Mapping Populations in Vegetable Brassicas Genetics

Genetic mapping requires a suitable molecular marker system, software for computational data analysis and specific types of plant mapping populations. There are several types of mapping populations that are widely used in genetic linkage mapping: F₂, backcrosses, doubled haploids (DHs), recombinant inbred lines (RILs) and near-isogenic lines (NILs). Of crucial importance for mapping populations development is selection of parents. The parents of the mapping population must show sufficient variation in the traits of interest at both the DNA sequence and phenotype level. The choice of appropriate mapping population must be made in view of the specific aim of genetic map construction and the types of genetic markers which are to be used. If one wants to map specific loci, short-term mapping populations like F₂ or backcrosses may be used, but if one wants to pursue a global genome mapping, long-term mapping, populations like DHs, NILs or RILs should be selected (for more details about the type of mapping populations see Hittalmani et al. 2002 and http://www.scribd.com/doc/6229849/Mapping-Population, Singh and Prasanna). The number of lines of mapping population is also very important and depends on the type of population and traits which are to be mapped. In genetic map construction for vegetable Brassicas many different mapping populations have been used (Table 4-4). According to enormous morphological diversity in the genus Brassica, there are many possibilities for parental selection. Intraspecific populations reveal high numbers of segregating loci like in B. oleracea but sometimes only in interspecific populations high polymorphism level is observed (Kianian and Quiros 1992a). Many details about the main existing mapping populations for the genus Brassica are available at http://www.brassica.info/resource/plants/mapping-populations.php.

4.5.2 Vegetable Brassicas Mapping Populations

4.5.2.1 F₂ Population

F₂ population usually is created by self-pollination or sib mating of F₁. F₂ lines are the products of single meiotic cycle and consequently, the ratio for the dominant marker is 3:1 and for the codominant marker it is 1:2:1. F₂ population can be developed relatively easily in a short time so it is a good population for preliminary mapping. This population cannot be used for quantitative traits loci (QTL) mapping because each individual is genetically different, which prevents their use in replicated trials over locations and years. Also exact replicas of lines cannot be obtained as it is a short-term population.
Table 4-4 Selected populations used in the Brassica genetic and breeding (based on http://www.brassica.info/CropStore/populations.php, with modifications).

<table>
<thead>
<tr>
<th>Species</th>
<th>Population name</th>
<th>Population type</th>
<th>Number of lines</th>
<th>Parental lines</th>
<th>Marker type</th>
<th>Distortion of marker segregation</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Horizon’ x ‘R500’</td>
<td>F&lt;sub&gt;2&lt;/sub&gt;</td>
<td>104</td>
<td>Canola-type cultivar “Horizon” and yellow sarson type “R500” (<em>B. rapa</em> ssp. <em>tricoloris</em>)</td>
<td>RFLP</td>
<td>85 loci with distorted segregation ratios, towards the “Horizon” allele</td>
<td>Chyi et al. 1992</td>
<td></td>
</tr>
<tr>
<td>Per x R500</td>
<td>F&lt;sub&gt;2&lt;/sub&gt; (F&lt;sub&gt;3&lt;/sub&gt;)</td>
<td>100</td>
<td><em>B. rapa</em> biennial cultivar Per and annual cultivar R500</td>
<td>RFLP</td>
<td>33 (23 %) loci with distorted segregation ratios; 30% of loci towards the “R500” allele, 40% of loci towards the Per allele</td>
<td>Teutonico and Osborn 1994</td>
<td></td>
</tr>
<tr>
<td>Per x R500</td>
<td>(F&lt;sub&gt;3&lt;/sub&gt;) RIL</td>
<td>87</td>
<td>Winter turnip rape (Per)(<em>B. rapa</em> ssp. <em>oleifera</em>) and a spring yellow sarson (R500)(<em>B. rapa</em> ssp. <em>tricoloris</em>)</td>
<td>RFLP</td>
<td>28 loci with distorted segregation ratios; 57 % of loci towards the „R500” allele</td>
<td>Kole et al. 1997</td>
<td></td>
</tr>
<tr>
<td>BraAGF2</td>
<td>F&lt;sub&gt;2&lt;/sub&gt;</td>
<td>94</td>
<td>Two DH lines of Brassica <em>rapa</em> ssp. <em>pekinesis</em>, Nou 7 (a clubroot-resistant line) and G004 (susceptible Chinese cabbage line)</td>
<td>RFLP SSR</td>
<td>15% of the total loci with distorted segregation rations</td>
<td>Suwabe et al. 2003; 2006; 2008</td>
<td></td>
</tr>
<tr>
<td>JWF3p</td>
<td>F&lt;sub&gt;2&lt;/sub&gt;</td>
<td>134</td>
<td>Developed from Chinese cabbage F&lt;sub&gt;1&lt;/sub&gt; cultivar Jangwon (<em>B. rapa</em> ssp. <em>pekinesis</em>)</td>
<td>RFLP SSR</td>
<td></td>
<td>Kim et al. 2006</td>
<td></td>
</tr>
<tr>
<td>Parent(s)</td>
<td>Generation</td>
<td>Number</td>
<td>Species/Accession</td>
<td>Technology</td>
<td>Distorted Segregation Rations</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>--------</td>
<td>-------------------</td>
<td>------------</td>
<td>-------------------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>R-o-18 x B162</td>
<td>F₂</td>
<td>114</td>
<td>B. rapa accessions B162 and Yellow-Sarson line (R-o-18)</td>
<td>AFLP SSR</td>
<td>16% of the total loci with distorted segregation rations</td>
<td>Soengas et al. 2007</td>
<td></td>
</tr>
<tr>
<td>BraCKDH</td>
<td>DH</td>
<td>78</td>
<td>Two different inbred lines of Chinese cabbage (B. rapa ssp. pekinensis), “Chiifu-401-42” and “Kenshin-402-43”</td>
<td>AFLP SSR RAPD ESTP STS CAPS</td>
<td>38.3% of the total loci with distorted segregation rations, generally towards male parent “Kensin” allele</td>
<td>Choi et al. 2007</td>
<td></td>
</tr>
<tr>
<td>B. nigra BniTIF2</td>
<td>F₂</td>
<td>83</td>
<td>B1164 selected from the Turkey population and B1157 selected from the India population</td>
<td>Isozymes RFLP RAPD</td>
<td>Towards one of parental alleles in the cluster</td>
<td>Truco and Quiros 1994</td>
<td></td>
</tr>
<tr>
<td>Catania x RC</td>
<td>BC</td>
<td>88</td>
<td>A rapid cycling accession (CrGC) and wild Italian accession (Catania)</td>
<td>RFLP</td>
<td>9% of the total loci with distorted segregation rations, half of the deviations were skewed towards a recurrent parent allele</td>
<td>Lagercrantz and Lydiate 1995; Lagercrantz 1998</td>
<td></td>
</tr>
<tr>
<td>B. oleracea</td>
<td>F₂</td>
<td>96</td>
<td>“Packman” broccoli (B. oleracea ssp. italica) and “Wisconsin Golden Acres” cabbage (B. oleracea ssp. capitata)</td>
<td>RFLP</td>
<td>9.7% of the total loci with distorted segregation rations</td>
<td>Slocum et al. 1990</td>
<td></td>
</tr>
<tr>
<td>C x CA</td>
<td>F₂</td>
<td>&gt;60</td>
<td>Collard (B. oleracea ssp. acephala) and cauliflower (B. oleracea ssp. botrytis)</td>
<td>RFLP</td>
<td>5.2% of the total loci with distorted segregation rations</td>
<td>Kianian and Quiros 1992a</td>
<td></td>
</tr>
<tr>
<td>C x B</td>
<td>F₂</td>
<td>&gt;60</td>
<td>Collard (B. oleracea ssp. acephala) and broccoli (B. oleracea ssp. italica)</td>
<td>RFLP</td>
<td>12.2% of the total loci with distorted segregation rations</td>
<td>Kianian and Quiros 1992a</td>
<td></td>
</tr>
<tr>
<td>U x CA</td>
<td>F₂</td>
<td>&gt;60</td>
<td>Wild kale from USSR (B. oleracea ssp. acephala) and cauliflower (B. oleracea ssp. botrytis)</td>
<td>RFLP</td>
<td></td>
<td>Kianian and Quiros 1992a</td>
<td></td>
</tr>
</tbody>
</table>

Table 4-4 contd...
<table>
<thead>
<tr>
<th>Species</th>
<th>Population name</th>
<th>Population type</th>
<th>Number of lines</th>
<th>Parental lines</th>
<th>Marker type</th>
<th>Distortion of marker segregation</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>K x I</td>
<td>F₂</td>
<td>52</td>
<td>Kohlrabi (B. oleracea ssp. gongylodes) and B. insularis</td>
<td>RFLP</td>
<td>59% of the total loci with distorted segregation ratios, towards maternal parent (kohlrabi) allele</td>
<td>Kianian and Quiros 1992a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F₂</td>
<td>120</td>
<td>cabbage line (No. 86-16-5) resistant to race 2 of Plasmodyophora brassicae (Woronin) and a rapid cycling line (CrGC No. 85)</td>
<td>RFLP QTL</td>
<td>Towards one or the other parental RFLP allele</td>
<td>Landry et al. 1992; Cheung et al. 1997</td>
<td></td>
</tr>
<tr>
<td>BolAGDH</td>
<td>BC₁</td>
<td>296</td>
<td>A12DHd (B. oleracea ssp. alboglabra) and GDDH33 (B. oleracea ssp. italica)</td>
<td>RFLP RAPD</td>
<td>7 loci with distorted segregation ratios</td>
<td>Ramsay et al. 1996</td>
<td></td>
</tr>
<tr>
<td>BolAGDH</td>
<td>DH</td>
<td>169</td>
<td>A12 (rapid cycling Chinese kale line; B. oleracea ssp. alboglabra) and GD (calabrese; B. oleracea ssp. italica)</td>
<td>RFLP</td>
<td></td>
<td>Bohuon et al. 1996</td>
<td></td>
</tr>
<tr>
<td>F₂</td>
<td>124</td>
<td></td>
<td>Cabbage (B. oleracea ssp. capitata) and broccoli (B. oleracea ssp. italica)</td>
<td>RFLP RAPD</td>
<td>22% of the total loci with distorted segregation ratios</td>
<td>Camargo et al. 1997</td>
<td></td>
</tr>
<tr>
<td>F₂</td>
<td>56</td>
<td></td>
<td>RCB (self compatible) and B. oleracea var. Green Comet</td>
<td>RFLP</td>
<td></td>
<td>Lan et al. 2000</td>
<td></td>
</tr>
<tr>
<td>F₂</td>
<td>247</td>
<td></td>
<td>RCB and B. oleracea var. Cantanese</td>
<td>RFLP</td>
<td></td>
<td>Lan et al. 2000</td>
<td></td>
</tr>
<tr>
<td>F₂</td>
<td>250</td>
<td></td>
<td>RCB and B. oleracea var. Pusa Katki</td>
<td>RFLP</td>
<td></td>
<td>Lan et al. 2000</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>Population Size</td>
<td>Breeders' Names and Varieties</td>
<td>Linkage Methods</td>
<td>References</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-----------------</td>
<td>------------------------------</td>
<td>-----------------</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F₂</td>
<td>246</td>
<td>RCB and <em>Brassica oleracea</em> var. Bugh Kana</td>
<td>RFLP</td>
<td>Lan et al. 2000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>97</td>
<td>DJ7032 (cauliflower line, Nedcha; <em>B. oleracea</em> ssp. <em>botrytis</em>) and DJ3753 (Brussels sprout hybrid Gower; <em>B. oleracea</em> ssp. <em>geminifera</em>)</td>
<td>AFLP</td>
<td>Sebastian et al. 2000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RFLP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Populations of this type were widely used in the initial mapping experiments for vegetables Brassicas. From amongst *B. rapa* Chinese cabbage cultivars, Michihili and Spring broccoli were selected as female and male parents, respectively to generate the mapping population (Song et al. 1991; Table 4-4). A set of 95 F$_2$ lines was analyzed to form a map-based on RFLP markers.

F$_2$ populations were often exploited to construct maps and to map genes or markers linked to agronomically important traits in *B. oleracea*. For example F$_2$ population composed of 96 individuals derived from a cross between “Packman” broccoli (*B. oleracea* var. *italica*) and “Wisconsin Golden Acres” cabbage (*B. oleracea* var. *capitata*) was used as a base for the first RFLP linkage map for *B. oleracea* (Slocum et al. 1990; Table 4-4). Four different F$_2$ populations were used for analysis of linkage arrangements and evolutionary implications in *B. oleracea* (Kianian and Quiros 1992a; Table 4-4). Three intraspecific hybrids were used to get F$_2$ populations: collard (B115) x cauliflower (B265), collard (B115) x broccoli (B008) and wild kale (B1661) x cauliflower (B265), each of the populations covered a set of 60 individuals. The fourth F$_2$ population, consisting of 52 lines was developed by interspecific cross between kohlrabi (B225) and *B. insularis* (B364). Combination of these populations was used for studying segregation of RFLP loci and map association. On this basis it was possible to observe distortions in the segregation ratios, gene distance as well as duplications of genes and whole genome allowing evolutionary implications. Sometimes F$_2$ populations are used for mapping after selfing, in the form of F$_3$ generations (F$_{2:3}$). F$_3$ population is used in particular applications such as mapping of recessive genes or sometimes QTLs, but like F$_2$ population, it is also not “immortal” (Singh and Prasanna http://www.scribd.com/doc/6229849/Mapping-Population). F$_3$ population can also be used when plants from F$_2$ generation are not available (e.g., when F$_2$ plants are too old). Such a conversion of F$_2$ population into F$_3$ generation was applied to chromosomal mapping in *B. oleracea* with expressed sequence tags (ESTs) from *A. thaliana* (Babula et al. 2003). F$_3$ generation was derived from F$_2$ population produced by the above-mentioned cross between collard (B115) and cauliflower (B265). The F$_3$ generation (originated from F$_2$ population developed from Chinese cabbage “Jangwon” *B. rapa* ssp. *pekimensis*) was also used to create sequence-tagged linkage map of *B. rapa* (Kim et al. 2006; Table 4-4).

4.5.2.2 Backcross Mapping Population

Backcross mapping population is obtained in the process, which begins with the cross between the chosen female and male parents. After obtaining F$_1$ generation, one single plant is selected to pollinate the plant used as a female parent. As a consequence, the first backcross generation comes into
existence \((BC_1)\). Backcross populations can be developed in a relatively short time although they are also not “immortal” populations. They are applied for marker-assisted breeding and in QTL mapping (Hittalmani et al. 2002; http://www.brassica.info/resource/plants/mapping-populations.php).

Backcross population was supposed to be promising for fine-mapping QTL in \(B. \text{oleracea}\) (Bohuon et al. 1996). Considering the presence of short fragment of chromosome introgressed into the recipient genome, QTLs may be mapped with very fine exactness. The starting point for the special population design, was a well known cross between two doubled haploid lines of \(B. \text{oleracea}\), A12DHd (A12) and GDDH33 (GD) (more details about the genetic cross between A12DHd and GDDH33 in Section 4.5.2.3; Bohuon et al. 1996; Table 4-4). A12 was used as a female parent (recipient of the fragment of chromosome) and GD was used as a donor of introgressed genetic material). One \(F_1\) individual was chosen to pollinate three A12 plants to obtain the first backcross generation (BC\(_1\)). In this backcross, BC\(_1\) population has a cytoplasm from A12. This process was repeated for the second backcross (BC\(_2\)) generation. The 176 individuals from BC\(_1\) were used for segregation analysis of RFLP and genetic map construction in order to reveal appropriate linkage between markers and identify the region of the genome with introgressed chromosomal fragment. The resultant map was compared with the previous maps such as those created by Bohuon and colleagues and particular individuals from BC\(_1\) were selected to be the parents of BC\(_2\). The approach with marker-assisted selection proved to be a successful way for generation of backcross mapping population (Ramsay et al. 1996). The same A12xGD backcross population was used for mapping QTLs for flowering time control in \(B. \text{oleracea}\) (Rae et al. 1999).

4.5.2.3 Doubled Haploids

Double haploids (DHs) are developed by culturing microspores or anthers. DHs result from one meiotic cycle similarly as \(F_2\) population, but the expected ratio for both types of markers is 1:1 (dominant and codominant). This is an “immortal” population because it can be replicated in different places and years without any genetic change. This stability guarantees the possibility of comparisons of the results obtained in different experiments. DH populations are ideal for QTL and global genome mapping (Hittalmani et al. 2002; Singh and Prasanna http://www.brassica.info/resource/plants/mapping-populations.php).

Embriogenetic nature of the \(Brassica\) microspores allowed a relatively rapid production of DH lines (Bohuon et al. 1996). With reference to many maps for \(B. \text{oleracea}\) based on \(F_2\) populations, DH mapping populations were indispensable for data comparison and integration. In \(Brassica\) and Oilseeds Research Department, John Innes Centre in Norwich, UK, a DH
population was produced for *B. oleracea*, which became one of the most widely used populations in the *B. oleracea* research, known as A12xGD or BolAGDH. A plant from DH lines of *B. oleracea* var. *alboglabra* (A12DHd) was pollinated by pollen from a single plant from the microspore-derived DH line of *B. oleracea* var. *italica* (GDDH33) to obtain a set of genetically identical F₁ offspring. From a cross between two highly polymorphic lines a set of 196 DH individuals was obtained. This population was used for comparative analysis of conserved fragments C genome between *B. oleracea* and *B. napus*, based on RFLP probes (Bohuon et al. 1996).

DH populations were used to make the first integrated linkage map for *B. oleracea* based on two different F₁-derived DH populations: the above-mentioned A12xGD and the second one named NxG. NxG (also known as BolNGDH) obtained by the cross between F₁-derived DH lines DJ7032 (*B. oleracea* var. *botrytis*) and DJ 3753 (*B. oleracea* var. *gemmafera*) (Sebastian et al. 2000; Table 4-4; See Section 4.7.3). The integrated map is much more saturated than the initial maps. The resultant map can be used for identification of homeologous loci and collinear chromosomal segments.

Two types of genetic markers were used for integration of data from two DH populations A12xGD and NxG. RFLP, AFLP and microsatellite markers were analyzed for both populations and after separate maps construction, integration was achieved by using JOINMAP software.

DH populations were frequently produced to study a particular feature, not the whole genome. A DH population was used for identification of two loci for resistance to clubroot in *B. rapa* (Suwabe et al. 2003; Table 4-4). Two DH lines were used as parents, “G004” and “A9709”. Line G004 was a clubroot resistance line developed by crossing “Homei P09” (a line with a high ability to regenerate from microsperm) and “Siloga S2” (a line derived by selfing of the clubroot resistant cultivar of European fodder turnip). The second parental line A9709 was clubroot-susceptible derived from a cross between two cultivar lines of Chinese cabbage: “Homei P09” and “Nozaki 2”. A cross between two parental lines “G004” and “A9709” gave an F₂ population, which was used for identification of the clubroot resistance loci.

DH populations were also used in map construction for oilseed crops: Indian mustard (*B. juncea*) and canola (*B. napus*). For *B. juncea* genome analysis, DH populations were obtained through a microspore culture from F₁ generation derived from the cross between Varuna and Heera (Pradhan et al. 2003). Parental lines differed significantly because they represented two distinct gene pools. Varuna was the crop that was largely widespread in India and Heera was the East European mustard line with canola quality. The DH population obtained was a set of 123 lines. It was used to produce high density genetic linkage map for Indian mustard based on RFLP and AFLP markers. The same DH population was used in the studies of QTL directly
related to the yield (Ramchiary et al. 2007). This population was also used in comparative mapping and evolutionary analysis of *B. juncea* and *A. thaliana* (Panjabi et al. 2008). Also, for *B. napus* many DH populations are available. One of the first DH populations for *B. napus* was constructed to map loci connected with controlling vernalization requirement and flowering time (Ferreira et al. 1994, 1995). TNDNDH population was obtained by crossing *B. napus* line Ningyou7 with Tapidor and then some F, individuals were used to obtain DH lines (Qiu et al. 2006). Another DH population for *B. napus* was produced by Parkin and colleagues (Parkin et al. 1995; 2005). It was the DH population from the cross between the winter breeding line CPB87/5 and the newly resynthesized *B. napus* line SYN1 (interspecific cross between *B. rapa* and *B. oleracea*). This population was used for identification of the A and C genomes of amphiploid genome *B. napus*.

### 4.5.2.4 Recombinant Inbred Lines

Recombinant inbred lines (RILs) can be produced in two ways: by prolonged self-pollination and by sib-mating of the offspring of individual members of F2 population to obtain fully homozygous lines (Broman 2005). After many rounds of crosses, RILs revealed the 1:1 genetic segregation ratio like DHs. RILs are obtained after many meiosis events so they are mosaics of parental genome, which makes them excellent for detection of closely linked markers. RILs have great potential in QTL analysis because of their genome structure and “immortal” character. They can be used over many seasons in different environmental conditions, which is significant for QTL mapping (Kole et al. 1997). For more details about RILs see Broman (2005) and Hittalmani et al. (2002).

The mapping populations of these types were used in construction of genetic linkage map for *B. rapa*. The F2 populations were derived from the cross between *B. rapa* cultivars, Per (a winter turnip rape) and R500 (a spring yellow sarson), which differed in many features (e.g., seed color, leaf pubescence, erucic acid content, vernalization requirement, flowering time, resistance to white rust, winter survival and freezing tolerance). RILs used for map construction were the F6 generation with high ratio of homozygosity (Kole et al. 1997; Table 4-4).

### 4.5.2.5 Near-Isogenic Lines

Near-Isogenic Lines (NILs) are generated by repeated self-pollination or by backcrossing of the F1 plants to the recurrent parents (Singh and Prasanna). NILs can be used to reveal particular gene effects and to display potential interactions between genes and environment. They are often used for QTL mapping (Yano and Sasaki 1997).
NILs are not frequently used for genetic studies in vegetable Brassicas like *B. oleracea* (Ping et al. 2007) or *B. rapa* (Muangprom and Osborn 2004). They were exploited for genome analysis in *B. napus* (Foisset et al. 1995; Li et al. 2006).

### 4.6 Mapping Software Used

The construction of genetic maps is based on the segregating data collected for experimental or natural populations. Together with the development of molecular techniques, the rapid increase in the number of loci has prompted development of specialized computer programs. Currently, different computer software packages are available for estimation of recombinant frequencies and map construction, among them the most popular are MapMaker (Lander et al. 1987; Lincoln et al. 1992) and JoinMap (Stam 1993). For the *Brassica* species most maps have been constructed using both the programs. The first maps of *Brassica* were developed by using the MapMaker program (for example Song et al. 1991; Chyi et al. 1992; Kole et al. 1997). However, construction of high-density and consensus maps, by consolidation of all segregating data derived from different mapping populations, required more advanced approaches, such as JoinMap, adapted by the majority of research groups (for example, Sebastian et al. 2000). Both programs use a similar approach for constructing maps and the same types of mapping populations: BC1, F2, RILs, doubled haploids, and the outbreeders full-sib family. They are based on efficient algorithms that allow simultaneous multipoint linkage analysis of numerous loci. Hence, the loci identification and order within the linkage groups obtained by these two programs are similar. The only major difference noted was the length of the maps. For example, Qi et al. (1996) reported that the length and observed genome coverage with loci mapped was greater when obtained with MAPMAKER than with JOINMAP, which resulted from different algorithms used in the map distance calculation. Furthermore, JoinMap allows integration of the segregating data from different populations into a single consensus map. Construction of the genetic map is performed in a few steps. Initially, linkage analysis is performed to detect which loci are linked. Linkage groups are assembled using two-point data. This is done by calculating all the pairwise linkage distances among markers in the data set prepared. The basic number of the linkage groups should be equivalent to the number of chromosomes in a given species. Next, the order of the markers within a linkage group is determined using three-point analysis. Three-point analysis is used to find the most likely locus orders within a linkage group. Determination of the locus ordering on the linkage map requires the use of criteria and algorithms to find the optimal sequence of loci. The criteria include the maximum likelihood (Lander et al. 1987), the
minimum sum of adjacent recombination fractions (SARF), the maximum sum of adjacent LOD scores (SALOD) (Liu and Knapp 1990), the minimum number of crossovers and the “least square locus order” (Stam 1993). Map distances in centiMorgans are calculated from recombination frequencies using the Haldane (Haldane 1919) or Kosambi (Kosambi 1944) mapping functions.

4.7 Map Construction-Map Types

Molecular marker-based genetic linkage maps in poorly characterized species such as Brassicas are fundamental tools for different research purposes. They are useful in genome structure analysis, studies of genomic relationship, diversity and genome evolution of the cultivated Brassica species, in identification of quantitative trait loci (QTL) and genetic markers closely linked to important traits such as disease resistance, abiotic stress, and morphological traits and in breeding programs based on marker-assisted selection (MAS). Another goal of map construction is to support the genome sequencing project. Hence, individual maps are often constructed with a specific goal in mind. The first molecular maps of the Brassica species were constructed in the 1980s. Currently, over 1,000 different molecular markers have been mapped onto at least 15 genetic maps of the Brassica genus derived from a wide range of crosses (Lim et al. 2007). The detailed description of the genetic maps available for diploid species of Brassica genus is given below and in Table 4-5.

4.7.1 Preliminary Maps

The preliminary genetic maps of Brassica were mainly based on anonymous sequences that allowed their rapid saturation. The first linkage groups of B. rapa were published by McGrath and Quiros (1991) (Table 4-5). The map was constructed on the basis of a set of RFLP probes located previously on the B. oleracea chromosomes (McGrath et al. 1990) to compare conservation of gene synteny between these species. The authors identified 49 loci using 31 RFLP probes and four isozymes, from which 31 loci were genetically linked to eight groups, covering 262 cM of the B. rapa map. This analysis showed extensive gene duplication; 35% of the probes used detected duplicated loci and 19% disclosed a single unique locus. Similar duplication degree of genome was obtained for B. oleracea (McGrath et al. 1990). The B. rapa map was constructed on the basis of the distinct subspecies cross, hence distorted loci, intraspecific gene duplications or deletions were observed (Table 4-4). These changes indicated the genome divergence within this species and could be assigned to morphological diversity within species. The first detailed maps of B. rapa were constructed by Song et al. (1991)
Table 4-5 Summary of the *Brassica* genetic maps.

<table>
<thead>
<tr>
<th>Species</th>
<th>Type of marker</th>
<th>Source of probes</th>
<th>Total length (cM)</th>
<th>Number of probes</th>
<th>Number of markers</th>
<th>Average number of loci per probe</th>
<th>Average interval between markers (cM)</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. rapa</em></td>
<td>RFLP Isozyme</td>
<td>Genomic clones isolated from <em>B. oleracea</em> and <em>B. napus</em>, cDNA clones from <em>B. napus</em> seedlings, napin seed storage protein gene (Crouch et al. 1983)</td>
<td>262</td>
<td>35</td>
<td>49</td>
<td>1.5</td>
<td>5.3</td>
<td>McGrath and Quiros 1991</td>
</tr>
<tr>
<td></td>
<td>RFLP</td>
<td>A <em>PstI</em> library of “Westar” (<em>B. napus</em> L.)</td>
<td>1,876</td>
<td>269</td>
<td>360</td>
<td>1.3</td>
<td>5.2</td>
<td>Chyi et al. 1992</td>
</tr>
<tr>
<td></td>
<td>RFLP</td>
<td>A <em>PstI</em> genomic DNA library and a cDNA library from the <em>B. napus</em> cultivar Westar; <em>EcoRI</em> genomic DNA library from <em>B. rapa</em> cultivar Tobin, cloned genes from <em>A. thaliana</em></td>
<td>1,785</td>
<td>116</td>
<td>139</td>
<td>1.2</td>
<td>13.5</td>
<td>Teutonico and Osborn 1994</td>
</tr>
<tr>
<td></td>
<td>RFLP</td>
<td>A <em>PstI</em> genomic DNA library and a cDNA library from <em>B. napus</em> cultivar Westar; <em>EcoRI</em> genomic DNA library from <em>B. rapa</em> cultivar Tobin (Teutonico and Osborn 1994); cloned genes from <em>B. napus</em> and <em>A. thaliana</em></td>
<td>890</td>
<td>102</td>
<td>144</td>
<td>1.4</td>
<td>6.0</td>
<td>Kole et al. 1997</td>
</tr>
<tr>
<td>Method</td>
<td>Description</td>
<td>Count 1</td>
<td>Count 2</td>
<td>Count 3</td>
<td>Count 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSR</td>
<td>ssr markers described early by Suwabe et al. (2002, 2004); RFLP (Harada et al. 1988); RAPD</td>
<td>1,006</td>
<td>262</td>
<td>3.7</td>
<td>Suwabe et al. 2006</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFLP</td>
<td>RAPD primers, single-copy genes from A. thaliana for SNPs and indels</td>
<td>743</td>
<td>273</td>
<td>3.6</td>
<td>Suwabe et al. 2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNP and/or indel markers</td>
<td>B. rapa EST clones from four different tissue-specific libraries; GST representing genes from A. thaliana chromosomes 4 and 5 flowering-time related genes; SSR primers (Suwabe et al. 2002; Lowe et al. 2004)</td>
<td>1,287</td>
<td>396</td>
<td>545</td>
<td>1.3</td>
<td>2.4</td>
<td>Kim et al. 2006</td>
<td></td>
</tr>
<tr>
<td>AFLP</td>
<td>AFLP with use EcoRI/MseI and, PstI/MseI enzymes; SSR, STS and ESTP were identify from BAC-end sequences of a KBH library of the inbred line &quot;Chiifu-401-42&quot; (Park et al. 2005); CAPS assay based on monomorphic products of STS and ESTP; RAPD primers from the Operon Technologies (USA) and WA (Japan)</td>
<td>1,182</td>
<td>556</td>
<td>2.8</td>
<td>Choi et al. 2007</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSR</td>
<td>SSr primers from Kim et al. 1999, Suwabe et al. 2002, Lowe et al. 2004 and AAFC Brassica/Arabidopsis Genomics Initiative</td>
<td>664</td>
<td>316</td>
<td>246</td>
<td>0.8</td>
<td>2.9</td>
<td>Soengas et al. 2007</td>
<td></td>
</tr>
</tbody>
</table>

Table 4-5 contd...
<table>
<thead>
<tr>
<th>Species</th>
<th>Type of marker</th>
<th>Source of probes</th>
<th>Total length (cM)</th>
<th>Number of probes</th>
<th>Number of markers</th>
<th>Average number of loci per probe</th>
<th>Average interval between markers (cM)</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. nigra</td>
<td>Isozymes</td>
<td>Genomic and cDNA clones isolated from the <em>B. napus</em> libraries, clones of known function, RAPD primers from the Operon Technologies (USA)</td>
<td>677</td>
<td>76</td>
<td>124</td>
<td>1.6</td>
<td>5.5</td>
<td>Truco and Quiros 1994</td>
</tr>
<tr>
<td></td>
<td>RFLP</td>
<td>Genomic <em>PstI</em> clones from <em>B. napus</em>, and <em>B. oleracea</em>, <em>Brassica</em> cDNA clones, <em>A. thaliana</em> cDNA clones</td>
<td>885</td>
<td>158</td>
<td>288</td>
<td>1.9</td>
<td>3.1</td>
<td>Lagercrantz and Lydiate 1995</td>
</tr>
<tr>
<td></td>
<td>RFLP</td>
<td>Genomic <em>PstI</em> clones from <em>B. napus</em>, <em>B. rapa</em> and <em>B. oleracea</em></td>
<td>778</td>
<td>158</td>
<td>288</td>
<td>1.9 (2.6)*</td>
<td>2.7</td>
<td>Lagercrantz and Lydiate 1996</td>
</tr>
<tr>
<td></td>
<td>RFLP</td>
<td><em>A. thaliana</em> DNA fragments: genomic <em>PstI</em> clones, anonymous cDNAs, genomic clones, YAC-end and gene sequences</td>
<td>751</td>
<td>160</td>
<td>284</td>
<td>1.8</td>
<td>2.6</td>
<td>Lagercrantz 1998</td>
</tr>
<tr>
<td>B. oleracea</td>
<td>RFLP</td>
<td>Genomic DNA clones derived from three different subspecies of <em>Brassica</em></td>
<td>820</td>
<td>197</td>
<td>258</td>
<td>1.3</td>
<td>3.5</td>
<td>Slocum et al. 1990</td>
</tr>
<tr>
<td></td>
<td>Phenotypic</td>
<td>Clones isolated from a <em>B. napus</em> cDNA library; genome-specific probes isolated from the <em>B. oleracea</em> and <em>B. napus</em> genomic libraries</td>
<td>747</td>
<td>62</td>
<td>108</td>
<td>1.7</td>
<td>6.9</td>
<td>Kianian and Quiros 1992a</td>
</tr>
<tr>
<td></td>
<td>Isozyme</td>
<td>Probes from a library <em>B. napus</em>: embryo cDNA clones, seedling-specific cDNA, cruciferin and napin cDNA</td>
<td>1,112</td>
<td>149</td>
<td>201</td>
<td>1.3</td>
<td>5.5</td>
<td>Landry et al. 1992</td>
</tr>
<tr>
<td>Method</td>
<td>Source Description</td>
<td>Band Count</td>
<td>Heterozygosity</td>
<td>Consistency</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>--------------------</td>
<td>------------</td>
<td>---------------</td>
<td>-------------</td>
<td>-----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFLP</td>
<td>cDNA clones from <em>B. napus</em>, genomic DNA from <em>B. oleracea</em></td>
<td>1,606</td>
<td>218</td>
<td>310</td>
<td>1.4</td>
<td>5.7 Cheung et al. 1997</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFLP</td>
<td>Libraries of cloned <em>Brassica</em> <em>Pst</em>I fragments (Sharpe et al. 1995)</td>
<td>875</td>
<td>167</td>
<td>303</td>
<td>1.8</td>
<td>2.9 Bohuon et al. 1996</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFLP</td>
<td><em>EcoR</em>I genomic cDNA library from <em>B. rapa</em> cv. Tobin (pTG clones); <em>Pst</em>I genomic DNA library of <em>B. napus</em> cv. Westar (pWG clones); <em>Pst</em>I genomic DNA library from <em>Brassica</em>; S-glycoprotein and S-locus related protein; phosphoglucose isomerase</td>
<td>747</td>
<td>138</td>
<td>138</td>
<td>1.0</td>
<td>5.4 Ramsay et al. 1996</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFLP</td>
<td><em>Pst</em>I genomic DNA library and cDNA library from <em>B. napus</em> cv. Westar (pWG clones) and <em>EcoR</em>I genomic cDNA library from <em>B. rapa</em> cv. Tobin (pTG clones)</td>
<td>921</td>
<td>125</td>
<td>159</td>
<td>5.6</td>
<td>Camargo et al. 1997</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFLP</td>
<td>cDNA clones from <em>B. oleracea</em> and <em>B. napus</em> library (Harada et al. 1988; from C. Quiros)</td>
<td>1,606</td>
<td>310</td>
<td></td>
<td>5.1</td>
<td>Cheung et al. 1997</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFLP</td>
<td>Genomic <em>B. oleracea</em> and <em>B. napus</em> libraries, cDNA library, gene specific clones</td>
<td>1,738</td>
<td>167</td>
<td></td>
<td>10.4</td>
<td>Hu et al. 1998</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4-5 cont'd...
<table>
<thead>
<tr>
<th>Species</th>
<th>Type of marker</th>
<th>Source of probes</th>
<th>Total length (cM)</th>
<th>Number of probes</th>
<th>Number of markers</th>
<th>Average number of loci per probe</th>
<th>Average interval between markers (cM)</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>RFLP</td>
<td>Brassica RFLP probes; <em>A. thaliana</em> genome clones</td>
<td>875</td>
<td>82</td>
<td>13.3</td>
<td>Bohuon et al. 1998</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFLP</td>
<td><em>Brassica</em> genomic DNA clones</td>
<td>889</td>
<td>433</td>
<td>3.3</td>
<td>Sebastian et al. 2000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFLP</td>
<td><em>Brassica</em> genomic DNA clones</td>
<td>893**</td>
<td>547</td>
<td>2.6</td>
<td>Sebastian et al. 2000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFLP</td>
<td><em>A. thaliana</em> EST and full cDNA clones</td>
<td>1806; 855</td>
<td>95; 160</td>
<td>212; 306</td>
<td>2.2; 1.9</td>
<td>8.5; 2.8</td>
<td>Babula et al. 2003; Kaczmarek et al. 2009</td>
<td></td>
</tr>
<tr>
<td>SRAP</td>
<td>Genomic and cDNA clones, broccoli BAC clones, <em>B. oleracea</em> genes</td>
<td>703</td>
<td>192</td>
<td>1257</td>
<td>0.6</td>
<td>Gao et al. 2007</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* including monomorphic bands.
** integrated map of *B. oleracea.*
and Chyi et al. (1992) (Table 4-5). Song et al. (1991) constructed the map of *B. rapa* containing 280 RFLP markers and covering a genetic distance of 1,850 cM. A more saturated map based on diversity of parental lines was constructed by Chyi et al. (1992) (Tables 4-4 and 4-5). This map included 360 RFLP markers distributed along 10 linkage groups and covering a total of 1,876 cM. Both Song et al. (1991) and Chyi et al. (1992) observed numerous sequence duplications, which were either dispersed in the whole *B. rapa* genome or arranged in the duplicated regions. Song et al. (1991) found three linkage groups involved in the most significant duplicate linkages and similar structural rearrangements among them. These structural rearrangements within duplicated blocks were responsible for evolution of the *B. rapa* genome. Although, the next map of *B. rapa* developed by Teutonico and Osborn (1994) included 139 RFLP loci only, it has been used for comparative analysis of *B. rapa*, *B. napus*, *B. oleracea* and *A. thaliana* (Table 4-5). The loci localized on this map included three simply inherited traits such as leaf hairs (pubescence), yellow seeds and low seed erucic acid. All these loci identified were mapped into the 10 main linkage groups and one additional small group, covering a genetic distance of 1,785 cM. Like the other authors in earlier reports, Teutonico and Osborn (1994) noticed a high level of loci duplication in *B. rapa*, forming small repeated chromosomal regions. The lack of larger duplicated regions could result from accumulation of local rearrangements after duplication or limited polymorphism for the used probes. The F2 population as the basis of this map was later developed to recombinant inbred (RI) population from the same F1 plant (Kole et al. 1997; Table 4-4). It permitted construction of the first framework map for *B. rapa*. In the initial stage, this map included 144 RFLP loci corresponding to genomic clones, cloned genes from *B. napus* and *A. thaliana* and three QTLs controlling seed color, leaf pubescence and resistance to white rust. They covered a genetic distance of 890 cM with an average distance between loci of 6 cM (Table 4-5). Comparison of the F2 and RI maps indicated a similar loci order along the linkage groups in these maps. Only some minor deviations in the loci order and interval distances in both maps were observed. Similarly as in the previous studies, the level of duplicated loci was high. It confirmed the complexity of the *B. rapa* genome structure.

Despite the importance of *B. nigra* as one of the progenitors of the tetraploid *B. juncea* and *B. carinata* species, only a few maps of this species have been constructed. The first of them was published by Truco and Quiros (1994). This map included 124 markers, and 122 of them were linked into 11 linkage groups with the total map distance of 677 cM and the average map density of 5.5 cM per marker (Table 4-5). This map provided information on the genome structure and complexity revealing a high degree of sequence duplication. Truco and Quiros found that 40% of the markers mapped were
duplicated. Many of them were organized in intragenomic chromosomal segments and others were dispersed over the whole genome or formed tandem arrangements. It confirmed the hypothesis about the polyploid origin of the Brassica species (Prakash and Hinata 1980).

Slocum et al. (1990) constructed the first detailed map of B. oleracea including 258 RFLP markers arranged into nine linkage groups and covering a genetic distance of 820 cM (Table 4-5). The majority of the RFLP markers detected were duplicated, and these duplicated copies were located on different linkage groups or organized in tandem duplications in the same linkage group. A comparison of the loci order along linkage groups indicated a high degree of their collinearity, although some rearrangements within these duplicated regions were observed. These data supported previous cytological data indicating existence of duplicated regions in the B. oleracea genome. Also preliminary molecular data from analysis of morphological, isozyme and RFLP markers in series of monosomic alien chromosome addition lines of B. oleracea indicated a high level of gene duplication (McGrath et al. 1990). Additionally, they detected some rearrangements such as insertion or deletion in the two parental accessions (“Packman” broccoli and “Wisconsin Golden Acres” cabbage) (Table 4-4). These data were confirmed in a wide range of the Brassica accessions. The rearrangements identified and chromosome restructuring have occurred during the B. oleracea evolution. They were considered as basic mechanism for gene evolution and differential gene expression (Ohno 1970). Next, using the same F₂ mapping population, Kennard et al. (1994) analyzed location and variation of 22 morphological traits such as leaf dimensions, heading habit, internode distance and flower development in the B. oleracea genome. Using the earlier mapped RFLP markers in this population Slocum et al. (1990) detected most of the major genes and many minor genes controlling the morphological traits analyzed. They also observed possible duplications of trait loci by comparing them with duplicated markers identified in the earlier analysis (Slocum et al. 1990). However, no preferential associations were observed between the loci controlling morphological traits and the pairs duplicated loci. Kianian and Quiros (1992a, 1992b) constructed a composite map for B. oleracea based on four different populations (see Section 4.7.3). These four component maps were analyzed for the location, organization, and variation of multigene families including sequences homologous to rRNA, napin, cruciferin, self-incompatibility, isocitrate lyase and malate synthase in the B. oleracea genome (Kianian and Quiros 1992b). The presence of the same loci allowed comparison of segregating data and indicated the chromosomal aberrations in these populations. In parallel, another map of B. oleracea based on a cross between the cabbage line resistant to race 2 of Plasmodiophora brassicae and a rapid cycling line was developed (Landry et al. 1992). This map consisted of 201 loci including
genes controlling leaf morphology and biennial flowering and covered a total genetic length of 1,112 cM. Furthermore, two dominant QTLs for resistance to race 2 of the clubroot disease were detected. This map was updated by Cheung et al. (1997) (Table 4-5). They added 136 marker loci together with the earlier mapped loci the map included 310 loci arranged into nine main linkage groups and four smaller ones covering a genetic distance of 1,606 cM. It covered the greatest distance with more markers when compared to those used in the earlier *B. oleracea* maps, providing additional information on the *B. oleracea* genome organization. This map was compared with the *B. napus* map revealing homoeology between C genomes of *B. oleracea* and *B. napus*.

### 4.7.2 High-Density Maps

Fast development of different marker systems has advanced construction of the high-density genetic maps. The maps have been constructed de novo or through the extension of the already published maps with the help of additional markers, as discussed below. The new markers provided a more accurate picture of the genome organization. These extended maps have become an excellent tool for marker-assisted breeding and accurate studies of genome evolution.

Presently, some high-density maps are available for *B. rapa*. The high-density map (JWF3) of *B. rapa* published by Kim et al. (2006) includes 545 markers assigned to 10 linkage groups covering a total map length of 1,287 cM with an average interval of 2.4 cM (Table 4-5). Most of the markers were detected by RFLP analysis using the sequenced expressed sequence tag (EST) probes derived from four different *B. rapa* tissue-specific libraries. Furthermore, five homologues of *FLC* (MADS-box flowering-time regulator) genes have been located on this map. In addition, 21 SSR markers have been mapped to assign each linkage group to the existing published A-genome linkage groups of *B. rapa* (Suwabe et al. 2002) and *B. napus* (Parkin et al. 1995; Lowe et al. 2004). Assuming the average marker density of 2.4 cM and 529 Mb Johnston et al. (2005) estimated that one marker is per Mb. The authors presented the pattern of intragenomic duplications in *B. rapa* (Kim et al. 2006). They found that many chromosomal regions had two or three copies, which supported the earlier discovery described above. Also, the detailed analysis of *FLC* gene indicated its duplication. In parallel, the SSR-based map of *B. rapa* including SSR, RFLP and RAPD markers was published (Suwabe et al. 2006; Table 4-5). This map included 262 markers with a genetic distance of 1,005.5 cM. The mapped SSR markers allowed assignment of the corresponding linkage groups in the *B. rapa*, *B. oleracea*, *B. napus* and *A. thaliana* maps. Earlier, genetic analysis of two loci controlling clubroot resistance (*Crr1* and *Crr2*) in the same F₂ population had been
carried out (Suwabe et al. 2003). They found close linkage between these loci and SSR markers which can be used in marker-assisted selection. Moreover, they discovered that the clubroot resistance was under oligogenic control. Hence, three QTLs were identified and mapped on this map. Another *B. rapa* map published by Soengas et al. (2007) was used for genetic analysis of QTLs controlling resistance to black rot caused by *Xanthomonas campestris pv. campestris* (Pammel) Dowson. This map was constructed on the basis of 223 AFLP and 23 SSR markers covering a total map distance of 664 cM (Table 4-5). The order and orientation of each linkage group were determined with respect to the previously published maps (see Kim et al. 1999 and Suwabe et al. 2004). Finally, the reference genetic map (CKDH) for *B. rapa* was constructed as support of the *Brassica rapa* Genome Sequencing Project (BrGSP, Choi et al. 2007; Fig. 4-1; Table 4-5). This map included the total of 556 markers distributed into nine linkage groups with a total length of 1,182 cM and an average distance of 2.1 cM between two loci. These linkage groups were identified by using common SSR markers with the reference map of *B. napus* (Parkin et al. 1995). Recently an attempt was made at construction of the *B. rapa* high-resolution genetic map via mapping more than 1,000 SSR markers derived from BAC-end sequences, ESTs and BACs (Hong et al. 2008).

Also, for *B. oleracea* some high-density maps based on different mapping populations and set of markers have been constructed (for review see Babula et al. 2007). Bohuon et al. (1996) constructed the *B. oleracea* map with 303 RFLP loci corresponding to 167 informative probes using a highly polymorphic DH population (Table 4-4). These loci were distributed in nine linkage groups corresponding to nine pairs of chromosomes and covered a genetic distance of 875 cM. From among these probes, as many as 54% detected more than one polymorphic locus, confirming the duplicated nature of *B. oleracea* genome. Each individual linkage group was oriented in relation to chromosome arms and numbered according to the corresponding linkage groups of *B. napus* (Parkin et al. 1995). Because of the use of the same set of RFLP probes for construction of linkage maps of *B. oleracea* and *B. napus*, the structure of C genomes could be compared in these species (See Section 4.10). The parental individuals of the mapping population derived by Bohuon et al. (1996) have been used to construct the library of recombinant backcross lines covering the whole genome (Ramsay et al. 1996). Linkage analyses in BC1 generation allowed the detection of 138 markers arranged into nine linkage groups and covering a genetic distance of 747 cM (Table 4-5). This work demonstrated the use of recombinant backcross lines to QTLs location and marker-based selection. Camargo et al. (1997) constructed the genetic map of *B. oleracea* based on RFLP and RAPD markers to localize a single locus controlling self-incompatibility (SI). The map made by Camargo et al. contained 159 loci and covered a genetic distance of 921 cM (Table 4-5).
The single SI locus was mapped into linkage group 2. Earlier, three linked loci for the same SLG6 probe had been detected, although no locus was assigned to SI (Kianian and Quiros 1992a). The presence of a single locus was confirmed by analysis of SI reaction within individuals of the population they studied by pollinating with the one parent.

Another method for construction of high-density maps is based on the enrichment of the published maps with additional markers. For example, it has been used by Saal et al. (2001), who added 31 SSR markers to the

Figure 4-1 Pattern of intragenomic duplication in *Brassica rapa* L. Marker loci are shown on the left of each linkage group (R1–R10). The duplicated loci and chromosomal segments are shown on the right of each linkage group. Each of 10 linkage groups is assigned to another color in gray scale (reprinted with permission from Kim et al. 2006).
existing genetic map of B. oleracea (Kianian and Quiros 1992a, Hu et al. 1998). Earlier, these SSR markers had been identified as C-genome specific in the B. napus genome. The presence of these SSR markers in the two species allowed testing the usefulness of rapeseed SSR markers in their mapping in B. oleracea.

The availability of the whole A. thaliana genome sequence permitted easier transfer of sequence information to the Brassica species. This knowledge was used for construction of the detailed linkage maps of the Brassica species. These maps were developed on the basis of RFLP markers derived from the A. thaliana ESTs or cDNAs corresponding to the known genes by Lan et al. (2000) and Babula et al. (2003). Comparative analysis of the maps made for the two species permitted defining the pattern of chromosomal collinearity between B. oleracea and A. thaliana. Babula et al. (2003) added 212 loci corresponding to the 95 A. thaliana EST probes to the existing B. oleracea map (Kianian and Quiros 1992a, Hu et al. 1998; Table 4-5). It covered a genetic distance of 1,806 cM with the average locus density of 9 cM. Lately, this map has been enriched by an additional group of loci corresponding to the A. thaliana EST and cDNA clones (Babula et al. 2006; Ziolkowski et al. 2006; Kaczmarek et al. 2009). Finally, this map included 306 loci corresponding to 160 A. thaliana probes and covering a genetic distance of 855 cM (Table 4-5). Now, this map is enriched by about over 40 SSR markers to integrate it with the reference map of B. oleracea (Sebastian et al. 2000, D. Babula-Skowronska et al., unpubl.). The highest density maps for B. oleracea were constructed by Sebastian et al. (2000) and Gao et al. (2007) (Table 4-5). The first of them represented a B. oleracea consensus map including 547 markers and covering a total map length of 893 cM (Table 4-5 and See Section 4.7.3). This map was constructed by integration of all segregating data from two different DH mapping populations: the Chinese kale x calabrese (BoLAGDH map) and cauliflower x Brussels (BoLNGDH map) (Table 4-4). The BoLAGDH map was adapted from the study by Bohuon et al. (1996) and enriched with AFLP and SSR markers. It included 433 markers covering a genetic distance of 889 cM with an average locus interval of 3.3 cM. However, the BoLNGDH map was constructed by G. King’s group. It comprised 223 markers covering a genetic distance of 831 cM with an average locus interval of 4.7 cM. Additionally, each linkage group was assigned by physical mapping to their respective chromosomes using FISH (Howell et al. 2002; Fig. 4-1). This map was adopted by other laboratories for more detailed studies. For example Ryder et al. (2001) conducted analysis of the B. oleracea genome organization within the linkage groups O3 and O6 and the corresponding sequences in the A. thaliana genome. The selected regions included markers of important genes such as sporophytic self-incompatibility locus S, MADS box transcription factors BoAPI and BoCAL. Comparative analysis of loci from the B. oleracea linkage
group O6 and the A. thaliana BAC map provided evidence of collinearity between these species and locus duplication through identification of two regions A and B. Almost all marker loci within region A of the linkage group O6 corresponded to single orthologous loci on the A. thaliana chromosome 1, however region B represented paralog of region A with inversion. It was consistent with those of Lan et al. (2000). On the other hand, no evidence of internal duplication within the region of linkage group O3 was found. Recently, the highest-density genetic map for B. oleracea was constructed by Gao et al. (2007) (Table 4-5). This map includes 1,257 markers distributed across nine linkage groups and its length is 703 cM. Each linkage group was assigned to their respective chromosomes using FISH and to chromosomes of A. thaliana. Also, on the basis of common SSR markers this map was assigned to the reference map of B. oleracea (Sebastian et al. 2000). Because of the presence of genes involved in glucosinolate pathway, resistance to cotyledon stage downy mildew and inflorescence development, it can be useful for marker-assistant selection. Particularly interesting can be the linkage groups: BoLG2 including BoGSL-ELONG associated with cotyledon stage downy mildew resistance and BoLG9 comprising the region with the AOP gene family members associated with side chain modification of aliphatic glucosinolates.

The first high-density RFLP map of B. nigra was constructed by Lagercrantz and Lydiate (1995) (Table 4-5). It was developed on the basis of the backcross generation F1 with a highly heterozygous parent (Table 4-4). Initially, the segregating data for female (recurrent parent) and male (F1) were scored separately. Finally, a combined male and female map was constructed. It comprised 288 loci in eight linkage groups with a total map length at 855 cM. This map provided important information about the male and female recombination frequencies and B genome structure indicating the intragenomic duplication and evolution and collinearity level between B. nigra and B. oleracea. Significant differences detected in the recombination frequencies between the parents were attributed to genetic effect. The data suggested that enhanced male and female recombination frequencies could be in proterinal regions and at centromeres of chromosomes, respectively. Comparison of marker loci along all linkage groups identified revealed the presence of triplicated chromosomal regions covering the whole B. nigra genome. Their presence confirmed the duplicated nature of the B. nigra genome. On the other hand, some translocations covering chromosomal segments were observed. The map was used to establish B genome structure and compare A, B and C genomes in Brassica (Lagercrantz and Lydiate 1996; see Section 4.10). The results pointed out that the whole B. nigra genome was represented by duplicated chromosomal segments. Eight segments were identified, each of which had three copies. This finding confirmed the hypothesis that modern Brassica species including B. nigra derived from a
hexaploid ancestor. Finally, this map was updated with an additional set of markers derived from the *A. thaliana* probes (Lagercrantz 1998; Table 4-5). Two hundred eighty four loci were added to the existing *B. nigra* map (Lagercrantz and Lydiate 1995) covering a total map length of 751 cM. Each linkage group with one exception contained the loci detected with probes from all five *A. thaliana* chromosomes. These studies allowed a more accurate definition of the genome architecture of *B. nigra*. As mentioned above, the complexity of the *B. nigra* genome of duplication was detected by Lagercrantz (1998), which supported the hypothesis about a hexaploid ancestor of the *Brassica* species. He suggested that the present-day *A. thaliana* genome structure represents the ancestor.

### 4.7.3 Composite and Integrated Maps

The composite maps were obtained by alignment of common markers localized on a few different maps. They could be treated as the first integrated maps. As shown above, individual maps were constructed for special purposes; moreover, they were developed on the basis of different marker systems and populations. This fact together with the highly duplicated nature of the *Brassica* genomes and the presence of more numbers of gene copies made it difficult to transfer information between the maps and opened the possibility of erroneous attribution of the corresponding linkage groups in different maps. Despite these difficulties, mapping with the help of multiple populations offers some definite advantages over the mapping based on a single population. In particular, a greater number of loci can be placed onto the map, which gives greater genomic coverage of the genome and permits more accurate estimation of the map distances. Finally, the common polymorphic loci in all studied crosses increase the chances of segregations in other populations. Hence, they can be used as anchor markers to align the corresponding linkage groups from several maps. Recently, Cmap program was adopted to compare and integrate the existing genetic maps of the *Brassica* species and also with the *A. thaliana* genetic map (Lim et al. 2007). The information contained within these maps can be further enriched when these maps are synthesized into a single consensus map to represent a given species.

The first of the *B. oleracea* composite maps was developed by alignment of four linkage maps (Kianian and Quiros 1992a). These component maps were constructed on the basis of three intraspecific (collard x cauliflower; collard x broccoli; wild kale x cauliflower) and one interspecific (kohlrabi x *B. insularis*) F$_2$ segregating populations (Tables 4-4 and 4-5). Alignment of common linkage groups of these maps by using shared loci allowed generation of a combined *B. oleracea* RFLP map. This map included 108 RFLP, isozyme and morphological loci distributed into 11 linkage groups.
and covering a total length of 747 cM. Eight of the eleven identified linkage groups were assigned to their respective chromosomes by alignment with the gene synteny groups of the *B. campestris-oleracea* addition lines (McGrath et al. 1990). Comparison of the loci order along individual linkage groups of these four maps indicated frequent rearrangements such as translocations and inversions in the *B. oleracea* genome, even at the subspecies level. Moreover, 56% of the probes mapped on the composite map detected duplicated sequences. Like in the earlier studies, these results indicated that 44% of the *B. oleracea* genome is duplicated, supporting the hypothesis about its paleopoliploidal origin (Quiros et al. 1987; McGrath et al. 1990). The duplicated loci were randomly dispersed or formed tandem arrangement such as gene families and duplicated chromosomal segments. One of these populations derived from a cross between collard and cauliflower was further developed by Hu et al. (1998), Saal et al. (2001) and Babula et al. (2003). Hu et al. (1998) added to the existing map 167 RFLP loci covering a genetic distance of 1,738 cM (Table 4-5). These authors also confirmed the duplicated nature of C genome of *B. oleracea* by revealing duplicated loci within and between chromosomes. This map was aligned with three other, previously published maps (Landry et al. 1992; Ramsay et al. 1996; Camargo et al. 1997). Each of these maps was constructed independently with different sets of probes and segregating populations. Using 1–5 common RFLP markers for each of the nine linkage groups, four maps were aligned. Accurate consolidation of all linkage groups was not possible, but some chromosomal regions were enriched with many markers. Only, six of the nine linkage groups of the Camargo’s et al. (1997) map were aligned with the base map. Further, a homology for four linkage groups in the base map and the map of Landry’s et al. (1992) was found. Five groups were aligned in the map of Ramsay et al. (1996) and that of Camargo et al. (1997). Each of these groups was assigned to their respective chromosomes using the map constructed with monosomic alien chromosome addition lines by McGrath et al. (1990). In general, the linear order of markers was maintained in all maps, but the distances between markers varied. However, some chromosomal rearrangements were observed. For example, a linkage group in the map of Landry et al. (1992) carrying a clubroot resistance and consisting of markers from two other linkage groups was rearranged. This composite map enriched by a large number of markers could be useful in further targeted research such as gene cloning. The consolidation of some genetic maps was continued by Lan et al. (2000). They constructed a comparative map of *B. oleracea* and *A. thaliana* using selected *A. thaliana* ESTs as probes (Table 4-5; see Section 4.11). In this study they used a composite *B. oleracea* map that was developed on the basis of mapping of the common *A. thaliana* ESTs in four populations. The unique loci mapped in each of the four populations were integrated on the basis of the closest common
flanking loci. With the use of 217 *A. thaliana* ESTs, 113 *Brassica* *Pst*I genomic clones, 23 *A. thaliana* cDNA clones and four cloned RAPD-PCR products, 367 polymorphic loci were detected. The “one to one correspondence” and “duplication” models developed in this study confirmed the earlier data concerning the duplicated nature of the *B. oleracea* genome. According to these models, 41% of RFLP loci detected were duplicated. The majority of them were organized in large intra-chromosomal segments. The duplicated segments were observed on the *B. oleracea* chromosomes 1, 2 and 9. Analysis of the arrangement along all linkage groups allowed displaying at least 22 rearrangement breakpoints differentiating *B. oleracea* homologs. Although some regions could be triplicated, these models did not support triplication of the *Brassica* genome.

Sebastian et al. (2000) worked out the first integrated map for *B. oleracea* based on segregating data from two distinct DH mapping populations and using 105 common markers (RFLP, AFLP and microsatellites) for the component BolAGDH and BolNGDH maps (Table 4-5). The consolidation of the segregating data of both maps allowed generation of a high-density map including 547 markers and covering a genetic distance of 893 cM, which is a 99% of the *B. oleracea* genome. The order and orientation of each linkage group was established with respect to original BolAGDH map (Bohuon et al. 1996). Because of the duplicated nature of the *B. oleracea* genome, banding patterns of the parental lines for the multiple-copy clones were checked, so that only homologous loci were compared. The correlation test concerning the position of the homologous markers in both component populations revealed close conservation of marker order with small rearrangements. In two linkage groups (LGs 1 and 5) the marker order was identical and 5 linkage groups (LGs 4, 6, 7, 8, 9) showed single inversion, whereas in the LG 3 and LG 2 two and several inversions were detected, respectively. Similarity of the marker order in the BolAGDH, BolNGDH and integrated maps means that the integrated map can be used as a *B. oleracea* consensus map. At present this map is further integrated with the map constructed by Babula et al. (2003) and updated by Kaczmarek et al. (2009) (D. Babula-Skowronska et al., unpubl). In this process, the SSR markers and low-copy *A. thaliana* ESTs were selected as anchor loci to assign the corresponding linkage groups from both maps.

### 4.8 Molecular Cytogenetic Maps

The genetic maps, even with high-density genes/markers are never complete. What is more, they present relative distances between markers only, based on the frequency of meiotic recombination. More accurate information about the physical distribution of genes on the chromosomes originates from the cytogenetic maps showing their approximate positions,
however they characterize low-density genes. Therefore, to get maximum information from the genetic, cytogenetic and physical maps for the genome sequencing projects, these maps must be integrated. The first step of this process is alignment of the linkage groups with karyotype (Howell et al. 2002; Fig. 4-1). It allows determination of a framework of sequence-anchored BAC clones covering the whole genome, determining of the orientation of linkage groups with respect to chromosome arms and comparison of genetic and physical distance of selected markers. Integration of cytogenetic and genetic maps was carried out for many plant species such as wheat, barley or A. thaliana (for example Koornneef and Van der Veen 1983; Künzel et al. 2000; Sandhu et al. 2001).

For the Brassica species, karyotypes and various genetic linkage maps have been established by different research groups. Identification and cytological characterization of individual chromosomes in Brassica is problematic because of the small size of chromosomes, specific-chromosome landmarks, as well as similar chromosome lengths and arms ratios. Nevertheless, some karyotypes of Brassica using in situ hybridization have been developed (Armstrong et al. 1998; Howell et al. 2002; Koo et al. 2004).

Connection of linkage groups of the existing B. rapa genetic maps with individual chromosomes is an important phase of the Multinational B. rapa Genome Sequencing Project (BrGSP). Because of duplicated nature of the B. rapa genome, the anchoring of sequence contigs on the physical map and reconstruction chromosomes sequence can be problematic. However, the availability of two distinct maps for this species should facilitate this process (Kim et al. 2006; Choi et al. 2007). The first of them, the high-density map of B. rapa constructed by Kim et al. (2006) has been initially attributed to the other existing maps. Using 21 SSR markers designated to A genome of B. rapa and B. napus (Parkin et al. 1995; Suwabe et al. 2002; Lowe et al. 2004), each linkage group of this map has been assigned to an accepted classification of the B. rapa linkage groups (R1–R10). Comparison of the relative length of each linkage group with the length of corresponding cytogenetic chromosome identified by Lim et al. (2005) has revealed compatibility between the linkage groups and the chromosomes, while the nine linkage groups of the reference genetic map for B. rapa (CKDH) have recently been assigned to chromosomes by using locus-specific ESTs to identify the corresponding BAC clones in KBrH library (Choi et al. 2007; Hong et al. 2008). These BACs identified used as FISH probes allowed integration of the genetic and physical maps of B. rapa ssp. pekinensis.

Truco and Quiros (1994) assigned the B. nigra linkage groups to chromosomes using two sets of alien addition lines, B. napus-B. nigra and B. oleracea-B. nigra (Chevre et al. 1991; Quiros et al. 1991; This 1992). Seven of the probes used for construction of the B. nigra genetic map were located on the chromosomes of the addition lines B. napus-B. nigra. Their location
allowed assignment of three linkage groups B1, B3 and B4 to chromosomes 1, 5 and 2, respectively.

In *B. oleracea*, the first efforts to assign linkage groups to chromosomes were taken up by Kianian and Quiros (1992a). They assigned eight of nine linkage groups from a *B. oleracea* composite map to chromosomes on the basis of the data reported by McGrath et al. (1990). More accurate integration of the cytogenetic and genetic maps was carried out by Howell et al. (2002). They selected the *B. oleracea* var. *alboglabra* line A12DHd, which was previously used as a parent of the mapping population for the construction of the reference genetic map (Bohuon et al. 1996; Sebastian et al. 2000), to produce partial karyotype (Armstrong et al. 1998) and a BAC library. Integration of both maps of the *Brassica* C genome was achieved with 22 probes representing 19 loci. They used genomic plasmids, cosmids and BAC clones with genetically mapped inserts as FISH probes for their localization on the cytogenetic map. Because of duplicated nature of the *B. oleracea* genome, it should be verified if individual FISH-based loci corresponded to the same locus on the genetic map. In the study by Howell et al. (2002) individual linkage groups of the *B. oleracea* genetic map were assigned to corresponding chromosomes of the karyotype. The assignment of linkage groups to chromosomes was as follows: Linkage group (LG) O1 to chromosome 8, LG O2 to chromosome 5, LG O3 to chromosome 1, LG O4 to chromosome 2, LG O5 to chromosome 6, LG O6 to chromosome 9, LG O7 to chromosome 4, LG O8 to chromosome 7 and LG O9 to chromosome 3 (Fig. 4-1). Additionally, eight of the nine linkage groups were oriented with respect to chromosome arms. Comparison of the orientation of linkage groups and chromosome arms in the genetic and cytogenetic maps, respectively, proved that four chromosomes (2, 4, 7 and 9) were in the same orientation as corresponding linkage groups, whereas the other four chromosomes (1, 3, 5 and 8) were in the opposite orientation towards these groups. The remaining chromosome 6 was probably also ordered in the opposite orientation. Using these data, Gao et al. (2007) aligned the constructed transcriptome map for *B. oleracea* based on cDNA-SRAP markers to C-genome linkage groups and chromosomes of *B. oleracea* and *B. napus*. A similar strategy of integration of genetic and physical maps of *B. oleracea* has been followed by Ziolkowski et al. (2006) (Fig. 4-2). These authors analyzed four rearrangement breakpoints by BAC-FISH and RFLP mapping of three *A. thaliana* chromosomal regions. For analysis, 95 *Arabidopsis* BACs representing chromosomal segments on the chromosome 1, 2 and 3 were selected. They used BAC contings as FISH probes to identify homologous regions on the *B. oleracea* chromosomes. Additionally, 36 *Arabidopsis* ESTs from the physical mapped segments were chosen to identify the corresponding loci on the *B. oleracea* genetic map. It allowed the assignment of particular linkage groups to individual
Figure 4-2 Assignment of the linkage groups (O1–O9) of the integrated genetic map to chromosomes (1–9) in *B. oleracea* by using the corresponding RFLP loci and FISH probes, respectively. The linkage groups are represented by vertical bars with marked location of the specific RFLP loci corresponding to FISH probes. The chromosomes are shown on the right of the linkage groups in relative lengths with marked location of the BAC, SS rDNA and 45S rDNA probes. Open circles represent centromeres. Chromosomes 1, 3, 5, 6 and 8 are inverted to follow the orientation of linkage groups (reprinted with permission from Howell et al. 2002).
Brassica chromosomes. Additionally, localization of analyzed group of A. thaliana ESTs on the B. oleracea reference map BolAGDH facilitated the identification of corresponding chromosomes of the basic karyotype (Howell et al. 2002).

4.9 Intragenomic Duplications

Results of many studies of the plant genome structures based on molecular markers and cytogenetic data indicated their ancient polyploid origin (for example Helentjaris et al. 1988; Kianian and Quiros 1992a; Shoemaker et al. 1992). One of the consequences is the presence of a great number of duplicated genes (Bennetzen 2000). The first mentions the duplicated nature of the Brassica genomes that appeared in the 1930s when it was postulated that the diploid species represent secondary polyploids derived from ancestral genomes with a fewer number of chromosomes (Catcheside 1934; Röbbelen 1960; Prakash and Hinata 1980). Further, more detailed studies based on RFLP mapping and comparative studies between the A. thaliana and Brassica species revealed that 40 to 50% of the loci were duplicated within Brassica genomes (McGrath et al. 1990; Slocum et al. 1990; Kianian and Quiros 1992a; Kowalski et al. 1994; Truco and Quiros 1994; Lagercrantz and Lydiate 1996; Lagercrantz 1998; Lan et al. 2000; Babula et al. 2003; Parkin et al. 2005). They could be classified as randomly dispersed single loci, linked-gene families or blocks duplicated loci in non-homologous chromosomes (Kianian and Quiros 1992a). Hu et al. (1998) classified duplications into two major groups: intra- and interchromosomal ones. The origin mechanism of loci duplication is yet unknown. It is suggested that randomly dispersed duplicated loci have arisen by reverse transcription of mRNA or heterogeneous nuclear RNA with subsequent reinsertion into the genome (Helentjaris et al. 1988). The linked-gene families can be a result of unequal crossing-over events. However, the segmental duplication is known to originate from the overlapping of reciprocal translocations followed by hybridization (Gottlieb 1983). Another way of loci duplication probably follows the formation of the interspecific aneuploids by combining homoeologous chromosomes of related genomes after hybridization and amphiploidization.

Analysis of available genetic maps of Brassica indicates a partial duplication of genomes forming complex arrangements of chromosomal segments present in at least three copies (Kowalski et al. 1994; Lagercrantz and Lydiate 1996). These triplicated chromosomal regions are associated with ancient polyploidization of the Brassica genomes. The reiterated segments are characterized by significant collinearity, although some rearrangements such as inversions, insertions and substitutions have observed (for example Lagercrantz 1998; Babula et al. 2003; Parkin et al.
The sequence-level studies of duplicated segments indicated that some gene copies were deleted within individual regions (O’Neill and Bancroft 2000; Rana et al. 2004). These modifications within genomes can lead to formation of novel phenotypic variations of traits among these species (Paterson et al. 2001). This observation has opened a new approach to phenotypic evolution of polyploidy in plants.

The duplicated structure of the \textit{B. rapa} A genome has been confirmed by some other authors. McGrath and Quiros (1991) found extensive gene duplication including 11 of 31 analyzed probes in the \textit{B. rapa} genome. Of the 11 probes, 3 duplicated loci were located on the linkage groups and the remaining duplicated copies were unlinked. Similar results had been obtained earlier for \textit{B. oleracea} by using the same set of probes (McGrath et al. 1990). Also, Song et al. (1991) observed extensive sequence duplication because almost 37% of probes detected duplicated sequences in \textit{B. rapa}. Many of them formed linked arrangements located on different linkage groups. On the other hand, some small rearrangements such as insertions, deletions and inversions within these reiterated segments were found. Analysis of a high-density map of \textit{B. rapa} with 545 loci allowed revealing a duplication pattern of this genome (Kim et al. 2006). In accordance with it, the majority of chromosomal regions were reiterated two to three times. This duplication pattern of \textit{B. rapa} genome has been confirmed by Choi et al. (2007) (Fig. 4-3). They found the duplicated region on R2 and R10 corresponding to the single region on the \textit{A. thaliana} chromosome 5. It is in agreement with the discoveries of Parkin’s et al. (2005) and Kim’s et al. (2006), who found that many chromosomal segments reiterated 2 or 3 times exhibiting extensive triplication of A genome (Rana et al. 2004). On the other hand, Yang et al. (2006) observed, that almost 88% of genes located around the duplicated \textit{FLC} regions were deleted.

Also, evidence has been found of the duplicated nature of the \textit{B. nigra} B genome. Three different types of arrangements have been identified: scattered through the genome, organized in tandem (gene families) and in blocks of duplicated loci conserved in more than one linkage group. Truco and Quiros (1994) detected segmental duplication located on the linkage groups B1 and B5. On the basis of a limited number of duplicated sequences, they established that the chromosome corresponding to the linkage group B1 could have resulted from an exchange of chromosomes B2 and B4, while, the chromosome corresponding to the linkage group B6 could have been derived from the chromosomes B3 and B5. Lagercrantz and Lydiate (1996) using the \textit{Brassica} RFLP probes suggested that the entire \textit{B. nigra} B genome consists of large, duplicated segments which reiterated three times. Graphically, the whole genome of \textit{B. nigra} is represented by eight sets of triplicated chromosomal segments with almost completely conserved loci order. The triplication has been confirmed by Lagercrantz
with additional 284 loci detected by the *A. thaliana* probes. These data support the hypothesis that *B. nigra* has descended from a hexaploid ancestor. Many of the collinear segments detected in the *B. nigra* genome were short because of high frequency of rearrangements. Nevertheless, some examples of long collinear segments have been found in this genome. One of them covered at least 40 cM of the top *A. thaliana* chromosome 5 and was present in three homoeologous copies distributed over three different linkage groups: G2, G5 and G8. Two of these segments were characterized by high collinearity, while the third one contained large inversion when compared to that of *A. thaliana*. Another example was the 40 cM segment of the *A. thaliana* chromosome 2, which corresponded to three homoeologous segments in *B. nigra*. One of these segment included 50 cM region on the *B. nigra* linkage
group G6 and the other two on the *B. nigra* linkage groups G1 and G8 were interrupted by segments from the other *A. thaliana* chromosomes. These data indicated that none of the *A. thaliana* chromosomes were significantly over- or underrepresented in duplicate homeologous copies in *B. nigra*. Sadowski and Quiros (1998) corroborated the duplicated nature of the *B. nigra* genome by analysis of gene members of the single-copy *A. thaliana* gene complex. This complex included six genes (including RPS2 gene) located on the chromosome 4 and they were expanded to 17 loci in the *B. nigra* genome and distributed over five linkage groups. The gene clusters with four to five members were found on the two *Brassica* linkage groups, B1 and B4 and on other partial smaller segments on B2, B3 and B8. Only gene complex on the *B. nigra* linkage group B1 was completely conserved relative to the corresponding gene cluster in *A. thaliana*. Physical analysis of this segment based on PFGE proved its extended size in the linkage group B1 of *B. nigra* compared with the size of this cluster in *A. thaliana* which is in agreement with the genome-size difference between these two species (Arumuganathan and Earle 1991).

Similarly, many studies indicated the duplicated nature of the *B. oleracea* C genome supporting the hypothesis about its secondary polyploidization (Prakash and Hinata 1980). As follows from RFLP analysis, at least 35% of the genes identified in this genome have been duplicated (Slocum et al. 1990; McGrath et al. 1990; Hu et al. 1998; Lan et al. 2000; Babula et al. 2003). These loci have been organized in tandem and segmental arrangements on all chromosomes. Detailed analysis of duplicated segments distribution in the *B. oleracea* genome indicated that at least some of them could be triplicated (Kowalski et al. 1994; Ziolkowski et al. 2006) or represented by two to five copies (Truco et al. 1996; Lan et al. 2000; Babula et al. 2003). Kianian and Quiros (1992a) analyzing the linkage map generated from three intra- and one interspecific populations, observed that the distances and order of markers within duplicated chromosomal segments survived. On the other hand, some rearrangements at the subspecies level were identified. The majority of these chromosomal modifications were represented by small inversions followed by translocations (Kianian and Quiros 1992a; Babula et al. 2003). In addition to these variations, the tandem duplications of individual genes were observed (Babula et al. 2003). The authors suggested that these rearrangements appeared as a consequence of genome evolution and variation in the chromosome number in the *Brassica* species.

The duplication level of the *Brassica* A, B and C genomes detected indicate their complexity as a result of polyploidization during their ancestry (Kianian and Quiros 1992a; Lagercrantz 1998). Within duplicated segments found in the genomes, numerous rearrangements were observed (Gustafsson et al. 1976; Quiros et al. 1988; Kianian 1990). Accumulation of different chromosome rearrangements led to renewed genome diploidization. Some
specific chromosome rearrangements have been found in the *Brassica* genomes. Among them, small inversions were the most frequent forms of aberration followed by chromosomal translocations (Ryder et al. 2001; Babula et al. 2003). In addition to inversion and translocations, deletions seem to be another important molding force of the *Brassica* genomes (Song et al. 1991; O’Neill and Bancroft 2000; Babula et al. 2003; Rana et al. 2004; Parkin et al. 2005). On the basis of the above data, the hypothetical models of the *Brassica* species evolution were proposed. One of the hypotheses known as the triplication theory indicated that the *Brassica* genomes were composed of three rearranged variants of an ancestral genome and descended from a common hexaploid ancestor (Lagercrantz and Lydiate 1996). On the basis of comparative genetic mapping between the *Arabidopsis* and *Brassica* species the allopolyploid nature of the putative hexaploid ancestor was proposed (Lagecrantz 1998). Lysak et al. (2005) using the physical comparative chromosome painting between *A. thaliana* and 21 Brassiceae and related species proposed that the ancestral Brassiceae genome became triplicated via allohexaploidy 7.9–14.6 Mya. These analyses indicated significance of the chromosome fusion and fissions in karyotype evolution of the Brassiceae. Other evidence supporting the triplication theory was provided by genetic/physical analysis of the selected chromosomal regions of *A. thaliana* in the *B. oleracea* genome (Ziolkowski et al. 2006). However, some authors did not confirm this theory because of identification of the *Brassica* gene copies in the number smaller or greater than three. Another hypothesis was proposed by Quiros’ group, who suggested amphiploidization as the origin of genomes of cultivated *Brassica*. This supposition would imply that the ancestral genome was composed of 5–7 chromosomes, which differentiated into several related genomes with similar chromosome numbers (Truco et al. 1996). Then, they diverged from one another by geographical isolation, leading to changes in chromosomal structure due to the accumulation of translations. Hybridization, amphiploidization and subsequent backcrossing of these genomes led to the appearance of the diploid *Brassica* species. Further, aneuploidy and secondary chromosomal rearrangements modified the genomes resulting in the present-day *Brassica* genomes of 8, 9 and 10 chromosomes in *B. nigra*, *B. oleracea* and *B. rapa*, respectively.

As all mapping experiments are based on polymorphism of genetic markers, the final pattern of duplication/triplication genome is incomplete and requires further verification. Nevertheless, the present knowledge of genomic relationship among diploid *Brassica* species brings us nearer to understanding the evolution of *Brassica* and gene transfer between them.
4.10 Homeology of A, B and C Genomes in the Brassica Genus

Three diploids and three amphidiploid genomes formed from them belong to the cultivated Brassica species; each with a characteristic chromosome number: A genome \((n = 10)\) in B. rapa, B genome \((n = 8)\) in B. nigra and C genome \((n = 9)\) in B. oleracea, AB genomes in B. juncea \((n = 18)\), AC genomes in B. napus \((n = 19)\) and BC genomes in B. carinata \((n = 17)\). Results of the phylogenetic studies of diploid and amphidiploid species revealed a relationship between the above species known as the U-triangle (U 1935; see Chapters 6 and 8 in this volume). This relationship was confirmed by more detailed investigation including cytogenetic study (Prakash and Hinata 1980); analyses of nuclear DNA content (Verma and Rees 1974), isozymes (Coulthart and Denford 1982; Quiros et al. 1987), proteins (Vaughan 1977), rRNA genes (Quiros et al. 1987) and chloroplast DNA (Erickson et al. 1983, Palmer et al. 1983); artificial resynthesis of hybrids (Ramanujam and Srinivasachar 1943; Mizushima 1950; Olsson 1960) and RFLP marker analysis (Song et al. 1988; Hosaka et al. 1990). Analysis of the cytological, biochemical and molecular data showed that the three diploid Brassica species evolved in two pathways: B. rapa and B. oleracea from one pathway having a common origin and B. nigra from another (Vaughan 1977; Attia and Robbelen 1986; Song et al. 1988; Warwick and Black 1991). The chromosomal collinearity level within the Brassica genus was determined by comparative genome mapping using common DNA probes (Slocum 1989; McGrath and Quiros 1991; Truco and Quiros 1994; Bohuon et al. 1996). Comparisons of gene order along chromosomes of the related species allow investigation and corroboration or rejection of the evolution hypotheses of their origin. The majority of comparative genome structure analyses among species of the Brassica genus concerned B. rapa and B. oleracea (A and C genomes, respectively), B. oleracea and B. napus (C and AC genomes, respectively), B. rapa and B. napus (A and AC genomes, respectively) and B. rapa, B. nigra and B. oleracea (A, B and C genomes, respectively) (Table 4-6). Nevertheless essential similarity of the diploid and amphidiploid Brassica genomes was proved, extensive rearrangements, forming unique combinations of chromosomal regions were observed as expected because of different chromosome numbers and genome evolution in the Brassica species. The studies have provided only a crude estimation of the genome structure and organization in the compared species because of insufficient marker saturation in the existing maps. Another problem was the duplicated nature of the Brassica genomes making the identification of orthologous regions in these species difficult.
Table 4-6 Characteristics of homoeology between A, B and C genomes of the *Brassica* species.

<table>
<thead>
<tr>
<th>Species</th>
<th>Genome homeology</th>
<th>Species</th>
<th>Common loci/probes</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. rapa</em></td>
<td>C genome</td>
<td><em>B. oleracea</em></td>
<td>95 cDNA probes</td>
<td>McGrath and Quiros 1991</td>
</tr>
<tr>
<td>(A genome)</td>
<td>A genome</td>
<td><em>B. napus</em></td>
<td></td>
<td>Hoenecke and Chyi 1991</td>
</tr>
<tr>
<td></td>
<td>A and C genomes</td>
<td><em>B. napus</em></td>
<td></td>
<td>Teutonico and Osborn 1994</td>
</tr>
<tr>
<td></td>
<td>B. rapa</td>
<td><em>B. oleracea</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. nigra</em></td>
<td>A and C genomes</td>
<td><em>B. rapa</em></td>
<td>116 SSR loci</td>
<td>Choi et al. 2007</td>
</tr>
<tr>
<td>(B genome)</td>
<td>B. napus</td>
<td><em>B. oleracea</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A genome</td>
<td><em>B. rapa</em></td>
<td></td>
<td>Suwabe et al. 2008</td>
</tr>
<tr>
<td><em>B. oleracea</em></td>
<td>A genome</td>
<td><em>B. rapa</em></td>
<td></td>
<td>Lagercrantz and Lydiate 1996</td>
</tr>
<tr>
<td>(C genome)</td>
<td>C genome</td>
<td><em>B. napus</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A, B and C</td>
<td><em>B. oleracea</em></td>
<td>129 RFLP loci</td>
<td>Bohuon et al. 1996</td>
</tr>
<tr>
<td></td>
<td>genomes</td>
<td><em>B. nigra</em></td>
<td></td>
<td>Sadowski et al. 1996</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>B. rapa</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A genome</td>
<td><em>B. rapa</em></td>
<td>117 RFLP probes 7 RAPD primers</td>
<td>Cheung et al. 1997</td>
</tr>
<tr>
<td></td>
<td>C genome</td>
<td><em>B. oleracea</em></td>
<td>22 SSR loci</td>
<td>Gao et al. 2007</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>B. napus</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>B. napus</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Analysis of A and C genomes and their relationships to other *Brassica* genomes has been a subject of many studies. The first genetic comparison of A and C genome structures in *B. rapa* and *B. oleracea*, respectively has been carried out by Slocum (1989) and McGrath and Quiros (1991) (Table 4-6). They compared the genetic maps based on random genomic and cDNA clones. McGrath and Quiros (1991) tested over 100 cDNA clones to recognize the corresponding DNA sequences in both species. About 95% of the probes detected at least one polymorphic locus in these genomes. Using 27 common marker probes, McGrath and Quiros (1991) indicated similar gene copy number forming the conserved block loci of both genomes. Their results as well as that of an earlier study (Slocum 1989) indicated that the same set of single copy loci was observed in both species. Three single loci were located on one linkage group in *B. rapa* and *B. oleracea*. On the other hand, effects of genetic differentiation in each of the genome were observed. The cytogenetic analysis of meiosis in pollen mother cells of *B. rapa* and *B. oleracea* (*B. alboglabra*) performed by Heneen et al. (1995) indicated differences in chromatin condensation. This differentiation was further confirmed by genetic mapping using isozyme and DNA-based markers to discriminate between these genomes. The results of all studies suggested that *B. rapa* and *B. oleracea* originated from a common ancestor, although chromosome reshuffling after speciation of both species has occurred. The similarity of A and C genome structures can allow their sexual recombination forming amphidiploid *B. napus*.

Other studies concerned a comparison of A and C genome structure in *B. rapa* and *B. oleracea*, respectively, with that of the corresponding genome of the amphidiploid *B. napus* species. On the basis of the interspecific hybridization and cytogenetic data, *B. oleracea* and *B. rapa* were proposed as progenitors of the amphidiploid *B. napus* (U 1935). The above thesis was confirmed by RFLP mapping distinguishing A and C genome specific markers. Starting from the 1990s, numerous comparisons of C genome in *B. oleracea* and *B. napus* and A genome in *B. rapa* and *B. napus* revealed their conservation (Lydiate et al. 1993; Teutonico and Osborn 1994) (Table 4-6). Teutonico and Osborn (1994) were the first to compare genetic maps of three species, i.e. *B. rapa*, *B. napus* and *B. oleracea*. The *B. rapa* map constructed by these authors was first compared with the previously published map for *B. oleracea* (Camargo 1994). The comparison revealed many regions with conserved loci order in both maps, confirming a relationship between these species, however, some rearrangements such as deletions or insertions were observed. These rearrangements could affect the distances between these loci in the way reflecting differences in the genomes size of *B. rapa* and *B. oleracea*. Further, this *B. rapa* map was compared with the previously published map for *B. napus* (Ferreira et al. 1994) (Table 4-6). All common RFLP loci with the exception of two were preserved, a common loci order
along the corresponding linkage groups in both maps. Similarly, the
distances between the common loci were different in these species. As the
*B. rapa* map has been constructed on the basis of the cloned genes mapped
in *A. thaliana*, a comparison of the two maps was possible. Bohuon et al.
(1996) compared the *B. oleracea* high density map constructed by them with
C genome linkage groups in the previously published *B. napus* map (Parkin
et al. 1995; Table 4-6). Although the two maps contained 132 common probes
detecting 270 polymorphic loci in *B. oleracea* and/or *B. napus*, a comparative
analysis was carried out by using 129 common RFLP loci. Comparison
of the loci order along the corresponding C genome linkage groups of
*B. oleracea* and *B. napus* revealed a collinear pattern for 123 of 129 common
loci. The rest of the six loci did not show compatibility due to the scoring
errors (*flowers* on O3/N13 and pW197dNP on N15), their presence in the
rearranged region in one of the parents of the *B. oleracea* cross (pW197f and
pR36e on O7) or the detection of different loci by pR116 on O3/N13. The
analysis of the loci arrangement along the corresponding linkage groups
on the *B. oleracea* and *B. napus* maps revealed a high level of collinearity,
indicating similarity of C genome structure in both species. The total length
of the collinearity region was 702 cM and 779 cM in *B. oleracea* and *B. napus*,
respectively. This observation has been confirmed by Cheung et al. (1997),
who using the same set of RFLP probes and RAPD and STS primers, also
indicated partial collinearity of C genome linkage groups in *B. oleracea* and
*B. napus* (Table 4-6). The highest level of conservation was found between
the *B. oleracea* linkage group 1 and the *B. napus* linkage group 3. These
authors estimated that 85% of the *B. oleracea* genome had homoeology with
the *B. napus* genome and 48% of the *B. napus* genome was derived from
*B. oleracea*. Comparative analysis of the *B. oleracea* and *B. napus* maps has
also performed by Ryder et al. (2001). They aligned the homoeologous C
genome linkage groups of *B. oleracea* and *B. napus* including O3 and N13
and O6 and N16, respectively (Parkin et al. 1995) (Table 6). The analyzed
region of the linkage group N16 was also homoeologous with A-genome
chromosomes N6 and N7 (Cavell et al. 1998). In spite of the collinearity of
C genomes of *B. oleracea* and *B. napus*, some major genome rearrangements
between these species were observed. Within the conserved regions simple
inversions or translocation including some loci were detected. Interestingly,
the rearrangements were more often found in C genome of *B. napus* than
*B. oleracea*. Similarly, comparative analysis of A genome in *B. rapa* and
*B. napus* has been made. The first studies based on comparison of A-genome
linkage groups of both species have been performed by Hoenecke and Chyi
(1991). They observed in general the conserved markers order for most
linkage groups in *B. rapa* and *B. napus*. Additionally, some minor and two
large rearrangements were identified. The rearrangements detected were a
consequence of interspecific hybridization of A and C genome progenitors.
On the other hand, a homoeology between A and C genomes has been found, confirming the results of earlier studies (McGrath and Quiros 1991; Teutonico and Osborn 1994).

Development and general availability of SSR markers facilitated comparison of A, B and C genome structures in the *Brassica* species. Gao et al. (2007) by using 22 common SSR markers aligned nine linkage groups BoLG1-BoLG9 of their *B. oleracea* map to equivalent linkage groups O1–O9 and the corresponding chromosomes in this species (Sebastian et al. 2000; Howell et al. 2002) and the *B. napus* linkage groups N11–N19 (Bohuon et al. 1996; Lowe et al. 2004; Piquemal et al. 2005) (Table 4-6). Additionally, 77 genomic SRAP markers detected and mapped on their map were included in the ultra-dense *B. napus* map constructed by Sun et al. (2007). However, on the basis of 116 common SSR markers, a reference genetic map of *B. rapa* ssp. *pekinensis* (Choi et al. 2007) was aligned to the reference A genome map of *B. napus* (Parkin et al. 1995). The SSR analysis indicated higher transferability between A and C genomes species than between this group and B genome species (Lowe et al. 2004). It confirmed the known evolutionary relationship between different *Brassica* species. This map based on 29 common SSR loci was aligned to A genome of *B. juncea* (Axelsson et al. 2000; Table 4-6). These studies indicated general conservation of the marker order along linkage groups representing A genome, that is R1 to R10, N1 to N10 and J01 to J10 in *B. rapa, B. napus* and *B. juncea*, respectively. However, this order was disturbed by several minor rearrangements reflecting differences in the genome structure in the species studied. Recently, Suwabe et al. (2008) constructed an integrated linkage map of A genome for *B. rapa* and *B. napus* using 44 common markers, 41 SSRs and 3 SNP/InDel (Table 4-6). They established that the order of the common markers along the linkage groups of both maps were conserved indicating a high level of A genome collinearity in these species. On the linkage groups A4, A5 and A10 small internal inversions were observed. The above results led to the conclusion that after integration of A and C ancestral genomes to the amphidiploid genome of *B. napus*; the integrity of their chromosomes was disturbed. The first comparative studies of the three diploid genomes were performed by the Lydiate’s and Quiros’s groups (Lagercrantz and Lydiate 1996; Truco et al. 1996). In the Lydiate’s group three genetic maps of *B. rapa* (Sharpe et al. 1995), *B. nigra* (Lagercrantz and Lydiate 1996) and *B. oleracea* (Bohuon et al. 1996) generated by the same set of probes were compared (Fig. 4-4; Table 4-6). This study indicated that the *Brassica* A, B and C genomes were very similar in spite of different chromosome numbers. The same authors found a high level of genomes conservation indicating the genome triplication as a general pattern of the *Brassica* genomes. It suggested that all diploid *Brassica* genomes inherited complete but rearranged duplicated segments of ancestral genomes. Moreover, the fission and fussion processes were
Figure 4-4: Alignment of the *B. nigra* linkage groups with the corresponding segments of the *B. rapa* and *B. oleracea* linkage groups. The vertical lines with the marker loci on the left and right represent the linkage groups. G1–G8, R1–R10 and O1–O9 correspond to the *B. nigra*, *B. rapa* and *B. oleracea* linkage groups, respectively. Lines drawn between segments indicate the common loci (reprinted with permission from Lagercrantz and Lydiate 1996).

indicated to take place during the genomes divergence and they were identified as responsible for changes in the chromosome number.

Truco et al. (1996) compared the order of loci along the linkage groups in the maps of *B. rapa* (McGrath and Quiros 1991), *B. nigra* (Truco and Quiros 1994), and *B. oleracea* (Kianian and Quiros 1992a) developed by a set of common probes (Table 4-6). The intergenomic conserved regions with extensive reordering among A, B and C genomes, were found. The highest homology was observed for A and C genomes, which confirmed their common origin. The results of this analysis suggested the presence of at least five ancestral chromosomes that could be considered as a foundation to build the present-day genomes by duplicating and reshuffling the existing information.

More detailed analyses of the diploid *Brassica* genomes structure were performed by Sadowski et al. (1996) (Table 4-6). These studies were limited to genetic and physical analysis of a cluster of five genes selected from a 15 kb segment on the *A. thaliana* chromosome 3 in three diploid *Brassica* species. This gene complex was used as a probe to identify common chromosome segments in *B. rapa*, *B. nigra* and *B. oleracea*. Three of the five probes displayed RFLPs in all three populations, whereas two of them segregated only in
B. nigra and B. rapa. Analysis of the hybridization patterns indicated that the number of the cluster members varies among species, only one of them was present in the single copy in all three genomes. The genes At1, At2 and At4 were represented by two or three loci in the *Brassica* genomes, mapping to three linkage groups of *B. oleracea* (C1, C6 and Cx) and to two in both *B. nigra* (B1 and B2) and *B. rapa* (A3 and A10), although not all genes were revealed due to the lack of polymorphism. Therefore, physical mapping based on pulse field gel electrophoresis (PFGE) was used in order to detect all five genes on the linkage groups A10, B2 and C1 of *B. rapa*, *B. nigra* and *B. oleracea*, respectively. The genetic and physical analyses indicated conservation of these chromosomal regions in *Brassica* in terms of general gene organization, relative to *A. thaliana*. Moreover, the additional, partial complexes in all three *Brassica* genomes were detected. The most complete secondary cluster was found in *B. nigra* containing four genes. These genes were linked to common sequences as genes coding the storage protein napin in all three genomes.

### 4.11 Genome Conservation between the *Brassica* and *A. thaliana* Species

Genome structure analyses indicate that many genomes at least partially have remained conserved during evolution (O’Brien et al. 1988; Whitkus et al. 1992; Bennetzen and Freeling 1993). The first studies of the genome structure revealed significant conservation of gene content and order among different plant species that diverged from a common ancestor more than 50 million years ago (Mya) (Crepet and Feldman 1991). Comparative mapping with the use of a common set of markers allows identification of homoeologous or homologous loci within collinear chromosomal regions in related species as well as rearrangements differentiating the genomes compared. Results of numerous studies indicated that mostly small rearrangements often took place in the plant genomes. For example, in rice and *Arabidopsis* 20% and 60% of genes identified were rearranged, respectively (Vision et al. 2000; Bennetzen and Ma 2003). Presently, these changes are mainly attributed to polyploidization and rediploidization processes, whose effects in the form of segmental duplication can be observed even in all diploid flowering plants (Kellogg and Bennetzen 2004). Hence, many genes are represented by some homologs, making it difficult to identify the accurate orthologs in the species compared. Knowledge of the genome conservation and ancient chromosomal rearrangements allows drawing conclusions about the evolution of different species. Therefore, a major challenge to plant biologists is to explain the mechanisms involved in the structural and functional evolution of duplicated genomes. Comparative genome analyses based on cross-hybridization of different plant genomes...
allow identification of sets of orthologous genes retaining identical gene order in some species. These studies have been initiated between closely related both diploid and polyploid species including the *Brassica* ones (Bonierbale et al. 1988; Helentjaris et al. 1988; Hulbert et al. 1990; Lagercrantz 1998; Babula et al. 2003).

The *Brassica* species represent a group of crops closely related to the model plant *A. thaliana*, whose genome sequence had been already determined. The phylogenetic analyses indicated that these species diverged from their common ancestor 17–18 Mya (Yang et al. 2006). Many detailed studies proved that the sequence homology between the *A. thaliana* genes and their homologues in *Brassica* is in the range from 75% to 90% in coding and < 70% in intronic regions (Cavell et al. 1998; Grant et al. 1998; Wroblewski et al. 2000; Quiros et al. 2001). Thanks to this high sequence homology, the *A. thaliana* genomic and cDNA clones could be used as probes on Southern blots to detect related sequences in the *Brassica* genomes. Their use gave new insight into the structure and evolution of their genomes. Analysis of chromosomal regions in *A. thaliana* and *Brassica* crops permitted detection of the collinear segments with a conserved gene order and a high degree of sequence conservation between these species (for example Sadowski et al. 1996; Lagercrantz et al. 1998; O’Neill and Bancroft 2000; Quiros et al. 2001; Babula et al. 2003; Gao et al. 2004; Ziolkowski et al. 2006). These studies showed that the *Brassica* genomes were a mosaic of *Arabidopsis*-like segments (Lagercrantz 1998; Babula et al. 2003). Moreover, many of the *Brassica* chromosomal regions were duplicated/triplicated with respect to the homologous segments of the *A. thaliana* genome (Lagercrantz 1998; Lan et al. 2000; Babula et al. 2003; Parkin et al. 2005). The existence of multiple chromosomal segments with partially conserved gene order in both genomes indicated that they could originate from the whole-genome duplication events in the *Arabidopsis* and *Brassica* lineages. The first evidence in support of this theory was found by mapping a small number of genes in both the *A. thaliana* and *Brassica* species (Kowalski et al. 1994; Teutonico and Osborn 1994). Kowalski et al. (1994) constructed the comparative map of *A. thaliana* and *B. oleracea* based on the previously mapped and newly selected sets of probes (Chang et al. 1988; Slocum et al. 1990). They detected 11 collinear chromosomal regions with a conserved loci order in the *A. thaliana* and *B. oleracea* maps covering 24.6% and 29.9% of the genome, respectively (Table 4-7). Analysis of distribution of these segments along the linkage groups pointed to triplication of the *B. oleracea* genome. Also, the duplication of the *A. thaliana* genome was suggested. Further and more detailed studies based on the genetic and physical mapping, permitted a more accurate definition of the degree of chromosome collinearity between *A. thaliana* and *Brassica* crops. Recently, in *B. rapa*, a reference high-density map was aligned to the *A. thaliana* genome
Table 4-7 Characteristics of conserved chromosomal regions in the *A. thaliana* and *Brassica* species.

<table>
<thead>
<tr>
<th>Species</th>
<th>Average length of conserved region (cM)</th>
<th>Number of syntenic groups</th>
<th>Number of common loci</th>
<th>Number of breakpoints</th>
<th>Type of rearrangements</th>
<th>Map coverage (%)</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. rapa</em></td>
<td>No data</td>
<td>29</td>
<td>153</td>
<td>No data</td>
<td>Deletions, duplications, inversions</td>
<td>29</td>
<td>Choi et al. 2007</td>
</tr>
<tr>
<td><em>B. oleracea</em></td>
<td>21.3</td>
<td>11</td>
<td>110</td>
<td>26</td>
<td>Translocations, inversions</td>
<td>29.9</td>
<td>Kowalski et al. 1994</td>
</tr>
<tr>
<td></td>
<td>7.1</td>
<td>34</td>
<td>273</td>
<td>No data</td>
<td>Translocations, Deletions, duplications</td>
<td>&gt; 28</td>
<td>Bohuon et al. 1998; Lukens et al. 2003</td>
</tr>
<tr>
<td></td>
<td>No data</td>
<td>24</td>
<td>186</td>
<td>19</td>
<td>Duplications, Inversions, Inversions, duplications</td>
<td>No data</td>
<td>Lan et al. 2000; Babula et al. 2003</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>42</td>
<td>212</td>
<td>24</td>
<td></td>
<td>46.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19.3</td>
<td>24</td>
<td>306</td>
<td>14-34</td>
<td>Translocations, duplications</td>
<td>54</td>
<td>Kaczmarek et al. 2009</td>
</tr>
<tr>
<td><em>B. nigra</em></td>
<td>8</td>
<td>83</td>
<td>284</td>
<td>79-112</td>
<td>Inversion, translocations/fusions</td>
<td>100</td>
<td>Lagercrantz 1998</td>
</tr>
</tbody>
</table>
These studies indicated that 78% of the mapped markers had sequence homology to the *Arabidopsis* genome. A detailed analysis of distribution of these markers along all linkage groups allowed detection of 139 homologous regions identifying 29 syntenic groups of markers between these species. These groups covered 344 cM, representing 29% of the *B. rapa* reference map (Table 4-7). Some of them revealed extensive genomic rearrangements as inversions, duplications, deletions and overlappings of many syntenic regions. Lan et al. (2000) and Babula et al. 2003 (updated by Kaczmarek et al. 2009) constructed a comparative map of *B. oleracea* and *A. thaliana* based on the mapping of *Arabidopsis* ESTs to characterize genome organization of both species (see Section 4.7.2). The EST-based maps are more informative and recognize homologous regions between species. The simple *Arabidopsis/Brassica* “one to one” and “duplication” models accepted by Lan et al. (2000) indicated that 57% of comparative loci have been found in syntenic segments between *Arabidopsis* and *B. oleracea* (Table 4-7). It corroborated extensive synteny between both species. Also, further studies of Babula et al. (2003) showed extensive conservation of many chromosomal segments in *Arabidopsis* and *B. oleracea*. They found 42 regions with conserved gene organization covering 61.6% and 46.5% of the *A. thaliana* and *B. oleracea* maps, respectively (Table 4-7). In some cases these conserved regions included entire chromosomal arms in *Arabidopsis*. The updated map based on 160 *A. thaliana* ESTs and full-length cDNA allowed identification of 24 collinearity regions covering over 54% of the *B. oleracea* genome. These results were confirmed by LineUp algorithm specifying statistical significance of chromosomal homology detected on the basis of the density and order of genes (Hampson et al. 2003). Furthermore, the detected blocks of collinearity between *A. thaliana* and *B. oleracea* were compared with the conserved blocks found under the Ancestral Karyotype developed by Schranz et al. (2006). Lan et al. (2000) and Babula et al. (2003), later Kaczmarek et al. (2009) detected that each of the conserved regions in *A. thaliana* was homologous to at least two regions in *B. oleracea* confirming a secondary polyploid origin of Brassica. Simultaneously, Lukens et al. (2003) on the basis of the literature data (Bohuon et al. 1996) defined the most closely related segments in *B. oleracea* and *A. thaliana*. They proposed the criteria for distinction of the orthologous from the paralogous loci based on a comparison of the position of the sequenced *Brassica* loci on a *B. oleracea* genetic map with positions of their putative orthologs within the *A. thaliana* genome. Using these criteria they identified 34 syntenic loci on a *A. thaliana* chromosome: O2 and At5, O4 and At2, O5 and At1, O8 and At1, O9 and At5, respectively, although seven of the nine *B. oleracea* linkage groups had collinear regions with more than one *A. thaliana* chromosome.
Rana et al. (2004) detected identical order genes in *B. rapa* and *B. oleracea*, which confirmed that the two genomes were very similar and had a triplicated structure with respect to that of *Arabidopsis*. This similarity was expected, because the lineages of *B. oleracea* and *B. rapa* diverged only by about 4 Mya. Similarly, the comparative map of *A. thaliana* and *B. nigra* constructed by Lagercrantz (1998) allowed identification of 41 short collinear regions, distributed along all black mustard linkage groups (Table 4-7). He also identified some large chromosomal segments with conserved loci order. Analysis of this comparative map confirmed triplication of the *B.nigra* genome, i.e. all collinear regions were represented by three copies in *B. nigra*, while it occurred in its single homologous in *A. thaliana*. Further, the collinearity of chromosomal regions in the *A. thaliana* and *Brassica* genomes was tested at the micro-level including selected gene groups or chromosomal regions in both species. An example is the genetic and physical mapping of gene complex in the diploid *Brassica* genomes originally defined in *A. thaliana* (Sadowski et al. 1996). Combination of the linkage analysis and physical mapping by PFGE proved useful in defining the gene cluster in *Brassica* and permitted identification of the genes present in duplicated regions and determination of the physical sizes of all clusters in three diploid *Brassica*. Five linked genes being in a single copy on the *Arabidopsis* chromosome 3 were used as a probe. This analysis revealed the presence of multiple copies of this gene complex in the three *Brassica* genomes, although the number of cluster members varied among species. Interestingly, each of the three *Brassica* genomes had only a single complete cluster the same as in *Arabidopsis* and the other showed a greater or smaller rearrangement. Furthermore, the size of the whole cluster was larger in the *Brassica* genomes due to the gene structure and the sizes of intergenic regions. Similarly, Sadowski and Quiros (1998) using the map constructed by Truco and Quiros (1994), analyzed genetic and physical organization of the six gene-cluster (including *RPS2* gene coding resistance to *Pseudomonas syringae*) from the *A. thaliana* chromosome 4 in *B. nigra*. The RFLP analysis indicated a simple hybridization pattern obtained with the *RPS2* probe revealing a single locus and complex patterns obtained with the remaining five probes. In total, 17 RFLP loci were identified in the *B. nigra* genome by the six *A. thaliana* probes distributed over five linkage groups. The gene cluster with four to five members were found on the two *B. nigra* linkage groups and other smaller segments on three linkage groups. Only a single almost complete cluster with a total conservation of the *A. thaliana* complex was found. The physical analysis of this conserved segment showed differences in the size of this cluster in *B. nigra* and *A. thaliana*, which is in agreement with the genome-size difference of these two species. A different strategy of combined linkage analysis and physical mapping has been used by Ryder et al. (2001). They studied the organization of two regions in the *B. oleracea*
Genome selected from the integrated linkage map and the corresponding sequences in the *A. thaliana* physical map. The first of them included 74.8 cM region of the shortest *B. oleracea* linkage group O6 with the sporophytic self-incompatibility locus *S* and two copies of the MADS box transcription factor gene *BoAP1*. The second of them included 30.4-cM region from the longest *B. oleracea* linkage group O3 with a copy of the MADS box transcription factor gene *BoCALa* (Sebastian et al. 2000). They identified two regions within the linkage group O6 with the locus duplication and collinearity with *A. thaliana*. The linear markers organization indicated good conservation with respect to the corresponding region of *A. thaliana*. One of these two regions showed a total collinearity with a single segment on the *A. thaliana* chromosome I. However, the latter region represented an inverted version of the first region. They also found evidence for local rearrangements within an overall pattern of collinearity. Organization of markers from the *B. oleracea* linkage group O3 indicated the lack of internal duplication. However, this linkage group contained sequences homologous to several regions of the *A. thaliana* genome i.e. the *A. thaliana* chromosome I, II, IV and V. However, there are markers common to this region and the linkage group O6. Because of imperfections of genetic mapping such as a limited level of polymorphism, the chromosomal collinearity in the *A. thaliana* and *Brassica* genomes was not estimated. Recently, Gao et al. (2007) compared sequences of the 11 *B. oleracea* genes and three BAC-end and 155 cDNA markers to the *A. thaliana* physical map. Their results were consistent with data reported by Li et al. (2003) and partially in agreement with the alignment of C genome of the *B. napus* and *A. thaliana* chromosomes. The differences in C genome structure of *B. oleracea* and *B. napus* might reflect chromosomal changes taking place during the allopolyploization process of *B. napus*. The above mentioned data have been supported by cytogenetic studies and sequence analysis as described in Chapter 6. Besides the chromosomal collinearity between the *A. thaliana* and the *Brassica* species, numerous rearrangements have been found (Table 4-7). Kowalski et al. (1994) identified at least 26 extensive chromosomal rearrangements as translocations and inversions that occurred in *A. thaliana* and *B. oleracea* since the divergence of both genomes. Lan et al. (2000) found at least 19 chromosomal structural rearrangements that differentiate both these species. Some of these chromosomal breakpoints seemed to be common in the *B. oleracea* and *Arabidopsis* genomes pointing to the chromosome organization of their common ancestor. As in the earlier studies, many rearrangements such as inversion and translocation/fusions within the collinear segments are observed in *B. nigra* and *A. thaliana*, which have occurred since the divergence of both species from a common ancestor. Lan et al detected 79 to 112 different rearrangements depending on the method for estimating the number of these rearrangements used.
Finally, they presented the reconstruction of an ancestral genome structure having determined a minimal number of rearrangements.

The genomic duplications observed in the Brassica species and accumulation of mutations such as deletions, insertions and inversions including more chromosomal regions or individual genes could lead to rapid changes in genome structure and the divergence of gene function increasing phenotypic plasticity in the Brassica species. It was found that in B. rapa 62% genes were conserved within homoeologous segments and had their counterpart in Arabidopsis. Similarly, in B. oleracea 68% genes were conserved. These genes may also show function conservation. The relationship between gene localization within collinearity regions in B. oleracea and A. thaliana and their function conservation has been confirmed by Babula et al. (2006). Initially, they analyzed chromosomal organization of selected genes and their homologues involved in the ethylene biosynthesis and signaling pathways in the B. oleracea genome. Twenty two loci with sequence homology to these genes were placed on the existing B. oleracea map. Most of them were located within the collinearity regions in B. oleracea and A. thaliana. Next, changes in gene expression in response to ozone stress for three ACS genes in B. oleracea and A. thaliana were studied. These analyses confirmed functional conservation of the genes in both species. In recent years many studies suggested that the lack of some genes within homoeologous segments could result from gene loss or insertion that can be the main mechanism responsible for these changes. The variety of functions of each copy of the gene can lead to nonfunctionalization (one of duplicated gene copies is silenced), subfunctionalization (both copies may become partially compromised), or neofunctionalization (one of duplicated gene copies may acquire a novel function).

4.12 Conclusions

The current progress in the Brassica genome analysis is mainly associated with the availability of informative genetic maps, knowledge of the whole Arabidopsis genome sequence and development of sensitive cytogenetic methods such as BAC-FISH. These findings allowed pre-determination of the genomic structure of Brassica indicating the genomes triplication and presence of many rearrangements as a consequence of poliploid origin. Moreover, analysis of markers order along linkage groups of three diploid Brassica species disclosed a high degree of chromosomal conservation between A, B and C genomes. It is important in light of the advanced project of B. rapa genome sequencing (BrGSP), because B. oleracea and B. nigra will become the main beneficiaries of this knowledge. The first step is the alignment of the reference maps of the Brassica species. It is known that each of these maps has been constructed on the basis of different
mapping populations, which means they are difficult to compare. Presently, this process is facilitated via mapping SSR sequences. The sequences are easily transferable between different species allowing identification of the corresponding sequences in various genomes. The first studies confirmed the usefulness of SSRs in alignment of some *Brassica* maps, which was discussed in this chapter. Furthermore, SSR programs currently developed in *Brassica* provide a vast number of microsatellite sequences. For example, for high-resolution mapping in *B. rapa*, BACs, ESTs and over 1,000 SSRs have been developed.

Construction of molecular maps is of great significance for increased breeding efficiency as it permits analysis of QTL detection, marker- or gene-assisted selection, gene flow and fingerprinting for clone identification of economically important genes. Furthermore, the maps facilitate detecting of a target region by searching public mapping database. Development of integrated maps, DNA and cDNA libraries and comprehensive databases is of importance for genome sequencing and structure and functional genomics. The studies of expression of duplicated genes in the *Brassica* species provide insight into the role of polyploidization in divergence of gene function leading to the phenotypic plasticity of *Brassica*.

**Acknowledgements**

This work was supported by a research grant from the Polish State Committee for Scientific Research (grant no. PBZ-MNiSW-2/3/2006/19 awarded to JS).

**References**


Ayele M, Haas BJ, Kumar N, Wu H, Xiao Y, Van Aken S, Utterback TR, Wortman JR, White OR,
Town CD (2005) Whole genome shotgun sequencing of Brassica oleracea and its application
mapping of Brassica oleracea based on ESTs from Arabidopsis thaliana: complexity of the
involved in biosynthesis and signalisation of ethylene in Brassica oleracea and Arabidopsis
thaliana: identification and genome comparative mapping of specific gene homologues.
Genome Mapping and Molecular Breeding in Plants, vol 5: Vegetables. Springer, Berlin,
Beckmann JS, Soller M (1990) Toward a unified approach to genetic mapping of eukaryotes
based on sequence tagged microsatellite sites. Biotechnology (NY) 8: 930–932.
Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of Arabidopsis.
Genomics 19: 137–144.
Bennetzen JL, Freeling M (1993) Grasses as a single genetic system: genome composition,
Bennetzen JL, Ma J (2003) The genetic collinearity of rice and other cereals based on genomic
Bohuon EJR, Keith DJ, Parkin IAP, Sharpe AG, Lagercrantz U, Lydiate DJ (1996) Alignment
of the conserved C genome of Brassica oleracea and Brassica napus. Theor Appl Genet 93:
833–839.
The association of flowering time quantitative trait loci with duplicated regions and
Bonierbale MW, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones
Univ of Wisconsin, Madison, WI, USA.
Camargo LEA, Savides L, Jung G, Nienhuis J, Osborn TC (1997) Location of the self-
and experimental characterization of physically clustered simple sequence repeats in
Catcheside DG (1934) The chromosomal relationships in the Swede and turnip groups of
Cavell AC, Lydiate DJ, Parkin IA, Dean C, Trick M (1998) Collinearity between a 30-centimorgan
segment of Arabidopsis thaliana chromosome 4 and duplicated regions within the Brassica
length polymorphism linkage map for Arabidopsis thaliana. Proc Natl Acad Sci USA 85:
6856–6860.
Cheung WY, Champagne G, Hubert N, Landry BS (1997) Comparison of the genetic maps of
Chevre AM, This P, Eber F, Deschamps M, Renard M, Delseny M, Quiros CF (1991)
Characterization of disomic addition lines Brassica napus—Brassica nigra by isozyme,
fatty acid, and RFLP markers. Theor Appl Genet 81: 43–49.


Lagercrantz U (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. Genetics 150: 1217–1228.


Genetics, Genomics and Breeding of Vegetable Brassicas


Molecular Linkage Maps: Strategies, Resources and Achievements


ABSTRACT

The term “complex traits” relates to any phenotype which escapes from classical Mendelian single-locus type of inheritance (recessive or dominant). Many important biological characters, from grain yield or pathogen resistance to complex human diseases as asthma or diabetes, result from a segregation of numerous quantitative trait loci (QTL). The fact that total genetic variation underlying those phenotypes is divided into portions between a large number of loci greatly complicates attempts to identify regions related to that trait. An addition of environmental effects and interactions between those genes completes the picture. This chapter is devoted to molecular mapping of complex traits in plants. First, we review commonly used approaches to analyze the nature of genetic variation for quantitative traits, from the 1920s, when their basis was established, until today. The following two sections illustrate successes achieved to date on identifying QTLs related to morphology, flowering time control, pathogen resistance, transformation and plant regeneration and nutrient composition in vegetable Brassicas. The last part focuses on recently introduced technologies such as single nucleotide polymorphism (SNP), diversity arrays technology (DArT) or expression microarrays, which have not been extensively used for analysis of QTLs in Brassica species, yet are methods of choice for future projects concerning heterosis. Finally, analysis of the genetic basis of crop domestication and adaptation along with epigenetic regulation of crop development are discussed.

Keywords: Brassica, QTL, molecular mapping, linkage disequilibrium, pathogen resistance, array-based technologies, crop domestication
5.1 Introduction

Genetic variation underlying quantitative phenotypes such as plant yield, pathogen resistance, etc. result from a segregation of numerous quantitative trait loci (QTL), each explaining a portion of the total variation, and whose expression is modified by interactions with other genes and the environment (Mackay 2001). The development of DNA markers enabled the construction of saturated genetic maps and localization of QTL for various phenotypes in various organisms (Glazier et al. 2002). Owing to their favorable genetic attributes, plants are used as model organisms for the study of quantitative traits. They are particularly suitable for high-resolution mapping and positional cloning. Much of the early work in this field was carried out by crossing the individuals of distinct phenotypes to maximize both QTL effects and DNA marker polymorphisms and constructing large experimental populations. Precise evaluation of mean phenotypic values called for statistically significant data obtained from replicated tests performed in different environmental conditions on either clonally reproduced genotypes or stable mapping populations such as recombinant inbred lines (RILs) or introgression lines (ILs). The creation of nearly isogenic lines (NILs) that differ only at a single region increased the precision of estimates of the effects of individual QTL or the interactions between QTLs (Paran and Zamir 2003). Although NILs can help to define the position of a QTL to a small region and thereby provide a shortlist of candidate genes for further investigation, the identity of the gene(s) underlying a QTL must be confirmed through the genetic and/or functional complementation and gene expression analysis (Paran and Zamir 2003). Two contrasting attributes of QTL analysis are: the relative simplicity of QTL mapping (to date ~ 1,750 papers have been published on QTL mapping in plants alone; www.ncbi.nlm.nih.gov; April 2010) versus the difficulty in assigning genetic locus to the QTL (until 2005 only 12 genes underlying QTL loci have been cloned) (Mackay 2001).

Until recently, the genetic bases of quantitative traits were described mainly using statistical methods of heritability analysis. In order to gain information on molecular aspects of what is now called genetic architecture of quantitative traits, the explanation of participation of individual genes in variation underlying those traits is needed. Such a detailed description requires the integration of classical genetics, evolutionary and ecological genetics, molecular population genetics, molecular genetics, and developmental biology, at a level of resolution that has not yet been attained (Mackay 2001).
5.1.1 Theoretical View on the Genetic Basis of Variation Underlying Quantitative Traits

While the number of genetically described loci is limited, even though 39 eukaryotic genome sequencing projects have been completed (www.ncbi.nlm.nih.gov, 8 April 2010), the issue remains “is it because mutations of minute effect escape from typical mutational screens?” (Mackay 2001). This question cannot be answered until the nature of genetic variation for quantitative traits is resolved. There were a number of theories on this matter.

First was a classic thesis called Fisher’s infinitesimal theory. His description of a process of adaptation assumed that all inherited traits are shaped by countless number of mutations with tiny effects (Fisher 1930; Farrall 2004).

The continuation of this model was proposed by Mather, who stated that there were two separate categories of genes: “oligogenes” characterized by Mendelian inheritance and “polygenes” which controlled quantitative traits. Mutations regarding the latter category were limited to alleles with subtle effects (Mather 1941; Mather and Harrison 1949; Mackay 2001; Barton and Keightley 2002).

With the expansion of QTL mapping experiments, new evidence appeared that contradicted both those theories. These studies have reported QTLs that explained only 20% of the observed variation in traits. As a consequence of these observations, Robertson and Reeve (1952) stated that there were no separate classes of genes, but what was called oligogenes were subject to a whole range of mutational effects from lethal effects to almost impossible to distinguish isoalleles with quantitative effects. The extension of this theory was known as the “candidate gene” theory, which proposed “segregation of above mentioned isoalleles was a source of quantitative variation in situations when it co-localized with loci at which major mutations affecting the trait occurred” (Mackay 1985; Robertson 1985). Both theories gave a recombination of a central role as supportive of genetic advances but the latter also put mutation in this position (Hill 1982).

Fisher’s theory has been modified several times (Kimura 1983; Orr 1998). The latter author proposed an exponential model of QTL effects. In other words, quantitative variation was controlled by the small number of loci with large effects and large number of loci with progressively smaller effects (Farrall 2004). This theory, in contrast to the infinitesimal model, gave hope to the approaches explaining the heritability of quantitative traits.

An observed phenotype, in respect to the quantitative traits discussed here, result from the expression of genes involved in developmental, physiological and/or biochemical pathways, which interact with each other and are under the influence of environmental factors. One way to investigate the function of these genes is the classical genetic approach—mutagenesis.
The multiple constraints of mutational tagging of QTL candidate genes were extensively discussed in excellent reviews by Nadeau and Frankel (2000) and Mackay (2001). In conclusion, it must be admitted, that although the value of this approach cannot be overestimated, however, it cannot fully replace QTL mapping in identification of a subset of loci affecting variation of the trait.

5.1.2 QTL Mapping

To perform QTL mapping one needs two things: a population in which phenotypic variation for the analyzed trait can be observed and DNA markers that detect polymorphisms in the experimental population, which then serve as loci for linkage map construction. The principle of QTL mapping, arguably traceable back to the 1920s (Sax 1923), is that simple traits or markers can be used to investigate the genetic control of more complex traits through linkage analysis. The basis of this approach has been further improved by findings of abundant DNA polymorphism most commonly small insertions/deletions and single nucleotide substitutions (Mackay 2001). Methods of detection of molecular variation have become simplified from the first generation of DNA markers (restriction fragment length polymorphism, RFLP, analyzed using Southern blots; Bonierbale et al. 1988) to high-throughput methods for both polymorphism discovery and genotyping (Kristensen et al. 2001; Gupta et al. 2008) (for more information see Chapter 4.2).

Improvement in mapping population development have also occurred (see Chapter 4.5). Traditionally, populations have been created by crossing two individuals with differences in trait phenotype. QTL mapping then utilized linkage disequilibrium (LD) between QTL alleles and marker alleles generated this way. The observation that LD was also present in many other types of populations (for details see Mackay 2001), have resulted in exploiting them in approaches known as association mapping (Gupta et al. 2005). This strategy has an advantage over linkage analysis, in respect to the most desirable attribute of QTL mapping - precision, which is directly proportional to the number of meioses sampled, correlated with the number of opportunities for recombination between the marker and the trait locus. Association studies, in contrast to artificially constructed populations with usually limited number of individuals or families, benefit from many generations of recombination, which have occurred in nature. Although, the level and distribution of LD has been investigated to the nucleotide level in many plant species (Lijavetzky et al. 2007; Mather et al. 2007; Xing et al. 2007; Fusari et al. 2008; Yan et al. 2009), *Arabidopsis thaliana* with it’s genome-wide studies is still the notable exception (Kim et al. 2007). Thus, QTL mapping usually includes: initial saturation of the genome with molecular...
markers performed using linkage analysis, followed by higher resolution confirmation studies of detected QTL, that end up with identification of candidate genes with recombination or association mapping.

5.1.2.1 Initial Genome Inspection

Key steps in the linkage mapping of QTL in inbreeding organisms were outlined in the previous paragraph. Along with the development of new methods of detecting polymorphism, there has been the improvement in statistical methods for mapping QTL and development of “golden rules” for experimental design and interpretation. The choice is between least-square (LS) (Soller et al. 1976) and maximum-likelihood (ML) (Lander and Botstein 1989; Knapp et al. 1990) methods taking into account both their advantages and disadvantages.

Many QTL mapping procedures prevent simultaneous evaluation of association of many QTLs with markers or marker intervals in the mapping population. To overcome possible favoritism on single marker and interval mapping analysis caused by the presence of multiple linked QTLs, one can use either composite interval mapping (CIM) (Jansen and Stam 1994; Zeng 1994) or the multiple interval mapping (MIM) (Kao et al. 1999) method. The former combines ML interval mapping with multiple regression, using marker cofactors to reduce the above mentioned bias and to increase the power to detect QTLs by decreasing the within marker-class phenotypic variation. The latter has been developed to provide estimates of positions and main and interaction effects of multiple QTLs. The investigator has to take into account that both methods are model dependent, and are influenced by different numbers of marker cofactors and window size (the region to either side of the test interval within which no marker cofactors are fitted) (Mackay 2001).

The need to improve production traits in agriculturally important crops has been addressed by mapping QTLs for traits of economic importance. The numbers of QTLs mapped in these studies are limited by sample size considered as a number of individuals in mapping population (correlated with the number of recombination events tested) and as a number of different alleles present in the two parents lines, which represent only a portion of the existing genetic variation. The methods to overcome this genetic limitation are four-way (Xu 1996) and eight-way cross design (Talbot et al. 1999; Mott et al. 2000), combining multiple line cross experiments (Xie et al. 1998) and utilizing parental strains derived by divergent artificial selection from a large base population (Nuzhdin et al. 1999).

Despite many obvious limitations, which lower the value of QTL mapping, this approach helps to understand the genetic basis of naturally occurring variation, albeit at the resolution of large chromosomal regions.
5.1.2.2 High-Resolution Mapping

High-resolution mapping aims to limit the size of the region exhibiting LD with the QTL to a single gene, by increasing recombination in the QTL region. Since there is a proportional relationship between the number of recombination and the number of generations, the required precision of mapping is achieved by exploiting further generations of the cross than F2 (Darvasi and Soller 1995). Also, the effects of single QTL can be magnified by constructing strains that differ only in respect to the region adjoining the QTL, from a whole chromosome (chromosome substitution lines) to a smaller interval (ILs and NILs; see Section 4.5 and 5.1). Numerous observations of trait phenotype performed on this type of lines increase the precision with which QTL genotype is specified.

It is gratifying to discover examples of effective fine-scale recombination mapping of QTL in plants, such as the localization of tomato fruit weight QTL, *fw2.2*, to a 1.6 cM interval containing four unique transcripts (Frary et al. 2000) and correspondence of tomato fruit-specific apoplastic invertase gene *Lin5* to the *Brix9-2-5* QTL, which controls fruit glucose and fructose contents, based on co-segregation of recombinants in a 484-bp region within this gene with QTL phenotypes (Fridman et al. 2000).

5.1.3 Gene Identification

Once a QTL is mapped to a region realistic for candidate gene identification, the question arises: “which one is it?” Again, the above mentioned studies on tomato can serve as an example of this approach. Observation of differences in QTL phenotype and polymorphic genotypes in a candidate gene for the same individuals was reported for tomato *Lin5* gene and a fruit sugar content (Fridman et al. 2000), provided evidence that the QTL was correlated with this gene. Likewise, functional complementation, in which the trait phenotype is recovered in transgenic organisms, can also help in gene identification. Functional tests can be carried out through overexpressing or down-regulating the gene(s) of interest (Salvi and Tuberosa 2005). These tests serve best in situations where QTL effects are large and their gene action is of a dominant type. Otherwise, it is necessary to assemble data from different sources, which together specify a candidate gene. Candidates can also be identified from genetic complementation tests, LD mapping, quantitative differences of gene expression and/or protein function, and exploring other species for orthologous QTL (Mackay 2001; Holland 2007). Expression profiling experiments may help in validation of candidate gene only when QTL for trait co-localize with QTL controlling expression of a candidate gene (Morgante and Salamini 2003).
5.1.3.1 Complementation Tests

One of the methods to demonstrate the identity of a causal gene and a QTL is the complementation test. This approach has specific requirements, which are: the possibility of preparing controlled crosses, the availability of mutants for all loci in the QTL region analyzed and clear discrimination of candidate genes. However, a weakness of this analysis is that failure to complement cannot discriminate between two types of genetic interaction: allelic and epistatic. Thus, final proof will include similarities in gene expression patterns (Doebley et al. 1997) and associations of molecular polymorphisms in the candidate gene with differences in the trait phenotype (Long et al. 1996, 1998).

5.1.3.2 Association Mapping

An introduction of association mapping method has been made in Section 5.1.1. It must be added here that besides an increase of precision of QTL mapping this approach has other advantages over linkage mapping in biparental offspring populations. It allows for detection of QTL in a wide range of germplasm rather than only those few that are polymorphic in classic populations. Since the population is natural, it takes less time and resources to assemble than standard mapping populations (Flint-Garcia et al. 2003). This method depends on measuring the linkage disequilibrium between QTL and other closely linked loci in a natural population (Mackay 2001). The extent of linkage disequilibrium is influenced by a number of factors including: the average amount of recombination per generation experienced by that region of the genome, the number of generations since the original mutation, and the population size (Hill and Robertson 1968; Falconer and Mackay 1996; Hartl and Clark 1997). The disadvantage of such collections or populations is that linkage disequilibrium also occurs for reasons other than genetic linkage, such as drift, selection, and admixture of populations. Association mapping has been applied to plants, at the level of both individual genes and the whole genome (Gupta et al. 2005; Mackay and Powell 2007).

Association mapping can be performed either to screen candidate loci in a previously defined interval or to look for associations without a priori linkage information. In a situation where genotypes at all polymorphic sites in the region of interest are known, it is possible to connect one of them with the sites generating the phenotypic effects (QTN-quantitative trait nucleotide) (Long et al. 1998; Long and Langley 1999). This is rarely true, and as a consequence association studies are conducted with only a subset of marker loci. As a corollary, the power to detect an association depends on the number of individuals sampled and the density of polymorphic markers.
5.1.3.3 Comparative QTL Analysis

There are several ways of exploiting synteny between closely related plant species in the course of QTL analysis. In a situation when one recognizes similar localizations of QTL for the same traits in different species, one can perform candidate gene verification in a model organism, and then look for linkage and association for the orthologous locus in other species. Conversely, high-resolution mapping could be carried out in one species and the correspondence of the genes in the interval with the QTL could be evaluated in a model organism (Mackay 2001).

There are several examples of some conservation of developmental pathways (flowering time control or plant architecture) among distantly or closely related plants species. QTL analysis of genes controlling photoperiodic response in rice and Arabidopsis has suggested their functional conservation, besides the obvious differences in the way they operate (reviewed in Paran and Zamir 2003).

5.1.4 Identification of Sequence Variation Underlying QTL

The obvious continuation of QTL analysis is identification of causal association between genetic polymorphism in gene sequences and phenotypic variation in traits analyzed (Mackay 2001; Morgante and Salamini 2003). And again, LD mapping is a method of choice. Studies of genotype-phenotype associations at candidate genes performed in humans and Drosophila revealed some interesting features. All significant associations have been found in introns and non-coding regions (for review see Mackay 2001), suggesting that variation in regulatory sequences can cause fine quantitative differences in phenotype (Morgante and Salamini 2003). In plants 12 genes corresponding to QTLs have been identified (Salvi and Tuberosa 2005). For two of them: fw2.2 in tomato (Cong et al. 2002; Nesbitt and Tanksley 2002) and tb1 in maize (Lukens and Doebley 2001; Hubbard et al. 2002), mutations have been detected in 2.6 kb and 50 kb upstream, respectively. The problems appear when it comes to describing the effects of mutations in non-coding regions, due to their distance from the regulated gene (Stam et al. 2002), laborious differentiation from neutral base pair substitutions and difficulty in predicting their phenotypic effect (Morgante and Salamini 2003). In conclusion, in quantitative control of the trait both the allelic diversity within gene transcription units and in distant regions contribute.

As a closing remark, it should be emphasized that understanding of molecular basis of phenotypic variation should not be restricted only to point mutations affecting genes but should also bear in mind the possible role of epistasis in QTL variation (Yedid and Bell 2002) and epigenetic
phenomena namely gene silencing, DNA methylation, RNA interference and heterochromatic DNA (Volpe et al. 2002; Zilberman et al. 2003). Also, it is worth mentioning that long terminal repeats retrotransposons have the ability to modify the expression of adjacent genes and can play a role in molecular basis of hybrid vigor (Kashkush et al. 2003).

5.2 Target Traits

5.2.1 Morphology

Plants from the genus Brassica have a long history of cultivation around the world as vegetable, oil, fodder, and condiment crops (for details see Chapters 1, 2 and 3). The wide variation in morphology evident today is likely to be in part due to differing selection pressure based on environmental growth conditions and the need of the consumer (Zhao et al. 2007). While QTL analysis has been used in other crop species primarily for disease resistance and flowering time control, the highly variable morphological forms of vegetable Brassicas provide additional opportunities for scientists who seek to understand the genetic basis of plant growth and development.

Many of the early studies observed morphological traits to be controlled by many genes (Pease 1926; Pelofske and Baggett 1979). One of the earliest attempts to identify genetic markers associated with genes affecting morphological and maturity characters was by Figdore et al. (1993; Table 5-1). Although their work focused mainly on the analysis of genetic basis of resistance to the clubroot pathogen, the trait loci controlling head type, color and extrusion were studied in a cross between susceptible cauliflower (B. oleracea var. botrytis; Early White) and resistant broccoli (B. oleracea var. italica; CR-7), using RFLP markers. For each of the traits analyzed, several chromosomal regions were found to contribute to each trait. This supported the conclusion that these traits are under polygenic control.

The complexity of genetic control of a subset of morphological traits was evaluated by Kennard et al. (1994; Table 5.1) in an F2 population developed from a cross between cabbage (B. oleracea var. capitata) and broccoli (B. oleracea var. italica). QTL analysis utilizing a single factor ANOVA enabled marker-locus associations to be identified for 16 of the 22 traits. In most cases, the same chromosomal region was detected to be correlated with different yet related traits. An example of this was the complex associations found among leaf dimension traits (i.e. lamina length, petiole length, basal petiole thickness and leaf length) on linkage group 4C, each of them having the same gene action (partial dominance), the same genotypic class having the largest mean (broccoli homozygous and heterozygous). Also, head forming leaf overlap, petiole length, lamina leaf, annual vs. biennial

Molecular Mapping of Complex Traits  205
habit were found to be highly associated on linkage group 5C. Previous studies, based on classical genetic methods, have observed heading habit, petiole length and lamina width to be linked (Pease 1926). Similar results were obtained for analysis of heading habit and biennial habit (Pelofske and Bagget 1979). As a corollary, this could be evidence that a common set of genetic loci control these morphological traits in different *B. oleracea* cultivars.

The degree of genetic divergence, as an effect of evolution of *Brassica* genus, has been elucidated by Lan and Paterson (2000; Table 5-1), in respect to curd-related traits using genotypes of different geographical origin: Italy, India and Thailand, which were Cantanese (CAN), Pusa Katki (PK) and Bugh Kana (BK), respectively. These three *B. oleracea* varieties, exhibiting different morphologies, were crossed to a common parent, the rapid-cycling Brassica (RCB), to create three *F₂* populations of *B. oleracea*. These populations were genotyped using RFLP markers and a total of 86 QTLs were identified for eight curd-related traits in *B. oleracea*. Using comparative mapping among the three mapping populations, they deduced that at least 67 different genetic loci were involved representing 54 ancestral genes (based on perceived patterns of genome duplication). These results supported previous research which found that the enlarged curd of cauliflower is under compound genetic control (Watts 1964; Kalloo and Bergh 1993; Kennard et al. 1994). Based on the observation of decreased complexity of genetic control of most traits, it was hypothesized that *B. oleracea* var. BK was a progenitor of cauliflower and broccoli. The comparison to the population used by Kennard et al. (1994; Table 5-1), who mapped similar traits, provided additional evidence on the differences in QTLs identified among these three populations. QTLs at nine locations were shared by both studies. Comparative mapping to *A. thaliana* identified several possible candidate genes for the *Brassica* QTLs.

The same RFLP-genotyped *B. oleracea* *F₂* populations were used by Lan and Paterson (2001; Table 5-1) to detect QTLs affecting the size of leaves and stems. They found 47 QTLs in three mapping populations, which represented approximately 35 different genetic loci, as inferred by comparison of the three linkage maps. There was some evidence that a proportion of these QTLs may have corresponded to homeologous locations in the *B. oleracea* genome. Mutants in potentially orthologous *Arabidopsis* genes equivalent to QTLs related to ancestral genes were identified.

Sebastian et al. (2002; Table 5-1), using a population of doubled haploid (DH) lines derived from a cross between two contrasting crop types, an annual cauliflower (*B. oleracea* var. *botrytis*) and biennial Brussels sprout (*B. oleracea* var. *gemmafera*), located QTLs controlling 27 morphological and developmental traits, including leaf, flowering, axillary bud and stem characters. The previously published study by Sebastian et al. (2000) reported
linkage map construction of *B. oleracea* genome utilizing the same population and RFLP, AFLP (amplified fragment length polymorphism) and simple sequence repeat (SSR) markers. QTL analysis based on multiple-marker regression identified 32 QTLs, affecting 17 morphological traits, up to four per trait. The range of additive genetic variation explained by them was between 6% and 43%. The work of Kennard et al. (1994; Table 5-1) compares closely to the results obtained in this study, concerning co-localization of markers associated with two leaf characters on linkage groups O6 and O7. Additionally, the similarity of mode of gene action provided evidence of common genetic control of these characters. Despite this, homology between those regions could not be evaluated due to incompatibility of marker sets used in both studies. In contrast, a QTL for petiole length found on O3 by Sebastian et al. (2002; Table 5-1) seemed to be homologous to a QTL detected in another *B. oleracea* study (Camargo et al. 1995).

Bettey et al. (2000; Table 5-1) investigated components of seed vigor and pre-emergence seedling growth traits in *B. oleracea* using a segregating mapping population of 105 DH lines. They found some correlation among traits, particularly among germination traits, and between seed weight and pre-emergence seedling growth traits. They concluded that germination and pre-emergence seedling growth were under distinct genetic control. QTL analysis revealed regions on linkage groups O1, O3, O6, O7 and O9 controlling these traits. Interestingly, two of these QTLs shared a similar position with QTLs for flowering time found previously by Bohuon et al. (1998; Table 5-1) in the same mapping population.

Wide variation for morphological traits exists also in *B. rapa* but less research has been conducted in this species compared to *B. oleracea* (see Chapters 2.3 and 3.3). A total of 22 traits including flowering time, seed and pod traits, growth-related traits, leaf morphology, and turnip formation was studied by Lou et al. (2007)(Fig. 5-1; Table 5-1) using several populations derived from three main morphotypes of *B. rapa* (the oilseed-, leafy-, and turnip morphotypes). QTL analysis was performed using both interval mapping and multiple-QTL model mapping methods. Altogether, 19 QTLs affecting morphological traits (excluding flowering time) were detected. Principal component analysis and co-localization of QTLs indicated that some of the loci controlling leaf and seed-related traits (on R07) and those for flowering time and turnip formation (on R02) might be the same.

Most of the above described QTL analyses concentrated on traits related to the morphology and development of shoot characters. The other group of vegetable Brassicas are plants grown for their succulent storage roots. One of the major vegetables within this group is turnip (*B. rapa* L. ssp. *rapifera* (Metzg.) Sinsk.). The first report on QTL analysis for root traits in Brassica crops was by Lu et al. (2008; Table 5-1) who detected 18 QTLs for three root traits: taproot thickness, length and weight. Simple interval
Figure 5-1 Locations of QTL for the traits analyzed in the four mapping populations of *Brassica rapa*. The linkage groups of different maps are aligned based on common SSR (S1–S10) or common AFLP markers (P Lou et al., unpublished data). The lengths of the arrows indicate the 2-LOD support intervals. The traits (see abbreviations in Table 5-1) are indicated above each column. The direction of the arrow’s head indicates the allelic effect: upward, RC-144 increases and CC-156 decreases for F2/3, YS-143 increases and PC-175/VT-115 decreases in DH-38 and DH-30; downward: CC-156 increases and RC-144 decreases, YS-143 decreases and PC-175/VT-115 increases in DH-38 and DH-30. The filling pattern of arrows refers to different groups of phenotypic traits. Flowering time; seed; plant height; leaf; turnip. S1a, BRMS096; S1b, Ra2G09; S1c, BRMS037; S2a, Na12H09; S2b, BrFLC2; S2c KS50030; S2d, BrMAF; S3a, BRMS043; S3b, BRMS042; S3c, BrFLC3; S5a, BRMS034; S5b, BRMS007; S5c, Ra3H10; S6a, BRMS014; S6b, Na12H07; S6c, KS51082; S7a, BRMS018; S7b, O112E03; S7c, Ra2A01; S7d, BRMS036; S8, Ra2E12; S9a, BRMS051; S9b, Na10A08; S9c, O112F02; S9d, O110D08; S10, BrFLC1. (reprinted from Lou et al. 2007 with permission of Oxford University Press).
Molecular Mapping of Complex Traits  209

mapping was conducted on an F2 population derived from a cross between two parents that have contrasting shoot and root characteristics: turnip \((B. \text{rapa} \text{ssp. rapifera})\) and non-headed Chinese cabbage \((B. \text{rapa} \text{ssp. chinensis})\). Results indicated that there is an extensive genetic variation for taproot thickness, length and weight traits in the analyzed hybrid population and that favorable alleles can be combined to improve storage-root yield.

5.2.2 Flowering Time Control

Morphological types of plant species of the genus \textit{Brassica} could be also differentiated based on their flowering habits: annual versus biennial. The former, represented by broccoli and cauliflower, groups plants which produce flowers during single growing season. Cabbage or kohlrabi exemplifies the latter, which will not flower unless exposed to low temperatures (vernalization). The recognition of genetic factors that control the switch from vegetative to reproductive stage will broaden the understanding of the basis of plant development (see Chapter 2.2.2.5). The modification of these factors will also enable the cultivation of a particular \textit{Brassica} species beyond their present geographical range.

In past decades, two contrasting concepts on genetic basis of control of flowering related traits have been proposed. On the one hand, single gene control has been put forward (Detjen 1926; Horovitz and Perlaska 1954; Walkoff 1963; Landry et al. 1992), on the other hand - potentially many genes have been suggested (Baggett and Wahlert 1975; Pelofske and Baggett 1979; Baggett and Kean 1989). The genetic control of flowering time was analyzed in an F2 population derived from a cabbage \((B. \text{oleracea} \text{var. capitata})\) by broccoli \((B. \text{oleracea} \text{var. italica})\) cross with the use of RFLP markers (Kennard et al. 1994). As a result, four regions of the genome controlling annual/biennial habit and flowering time were identified, and epistasis between two genes controlling flowering time was detected.

This approach was improved upon by Camargo and Osborn (1996; Table 5-1), who used in their study replicated F3 families, derived from a cross between the same \textit{B. oleracea} subspecies as Kennard et al. (1994; Table 5-1), but using different parental genotypes; this provided a better estimate of the genotype effects. Marker interval analysis indicated that flowering time was controlled by three or more loci, supporting other studies that reported oligogenic control. Two QTLs located on separate linkage groups were associated both with annual/biennial habit and flowering time. Comparison with previous analysis revealed the linkage of loci on LG2 and LG6 with a gene encoding S-glycoprotein, involved in self-incompatibility mechanism of \textit{B. oleracea} (Nasrallah et al. 1985) and petiole length affecting QTL (Camargo et al. 1995), respectively. A third region on LG 8 was associated with flowering time only. Consistent with
<table>
<thead>
<tr>
<th>Trait</th>
<th>Plant Material</th>
<th>Spatial Conditions</th>
<th>Temporal Conditions</th>
<th>Marker</th>
<th>Statistical approach</th>
<th>Software</th>
<th>QTL</th>
<th>Variation explained (%)</th>
<th>QTL position LG/Flanking markers/ distances (cM)</th>
<th>Authors</th>
</tr>
</thead>
</table>
| Resistance to race 2 of *Plasmodiophora brassicae*  
Leaf morphology (LMOR) | 78 F, *B. oleracea* var. *capitata* resistant No. 86-16-5 x rapid cycling *B. oleracea* CrGC No. 85 | growth cabinet | 43 days after inoculation | RFLP | Interval mapping | Mapmaker/ QTL | CR2A CR2B LMOR | 58 | 15 | LG6, 2NF11-2ND3 (22); LG1, 3NE4a-3ND3 (12); LG3, 3ND12, 21 | Landry et al. (1992) |
| Resistance to race 7 of *Plasmodiophora brassicae*  
(R) | 180 F, clubroot resistant broccoli *B. oleracea* var. *italica* CR7 x susceptible cauliflower *B. oleracea* var. *botrytis* | greenhouse | 35 days after inoculation | RFLP | Chi-square, R-square, F-tests | | | | | LG1, 14a; LG1, 34-75a (35); LG4, 48; LG9, 177b | Figdore et al. (1993) |
| Head type (H) | susceptible cauliflower *B. oleracea* var. *botrytis* “Early White” | | | | | | | | | LG9, 2; LG3, 96a-188(4); LG9, 2-37a(19) | |
| Head color (C) | | | | | | | | | | LG3, 96a-188(4); LG4, 71-59(25); LG9, 79-37a(39) | |
| Head extrusion (E) | | | | | | | | | | LG8, 6a-82a(6); LG1, 64 | |
| Heading data (maturity-M) | | | | | | | | | | LG1, 19-14a(72); LG4, 71-59(25); LG5, 210b-118(40); LG6, 110 | |

Table 5-1 An overview of QTL analysis for vegetable Brassica crops.
<table>
<thead>
<tr>
<th>Developmental traits:</th>
<th>Flora traits:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flowers</td>
<td>Annual/biennial</td>
</tr>
<tr>
<td>Days to bud to</td>
<td>Days from bud to</td>
</tr>
<tr>
<td>Flower</td>
<td>Flower</td>
</tr>
<tr>
<td>Size of cluster</td>
<td></td>
</tr>
<tr>
<td>Leaf traits:</td>
<td></td>
</tr>
<tr>
<td>Lamina length</td>
<td>Petiole length</td>
</tr>
<tr>
<td>Lamina width</td>
<td>Basal petiole width</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Days to bud</th>
<th>70–90 F</th>
<th>2 cabbage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wisconsin</td>
<td>Golden Acre</td>
<td>x Broccoli</td>
</tr>
<tr>
<td>Packman</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ELF</th>
<th>Single factor ANOVA, Chi-square tests, Pearson correlation coefficients</th>
<th>SAS GLM procedure, SAS FREQ, SAS UNI-VARIATE, SAS CORR</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.7–42.7</td>
<td>(single marker loci), 60.1</td>
<td>-</td>
</tr>
</tbody>
</table>

 Kennard et al. (1994)
<table>
<thead>
<tr>
<th>Trait</th>
<th>Plant Material</th>
<th>Spatial Conditions</th>
<th>Temporal Conditions</th>
<th>Marker</th>
<th>Statistical approach</th>
<th>Software</th>
<th>QTL Variation explained (%)</th>
<th>QTL position LG/Flanking markers/distance (cM)</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basal petiole thickness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaf length</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lamina length/width</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lamina/petiole length</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stem traits:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance between second</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and third node</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance between third</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and fourth node</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance of twenty</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>internodes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flowering time (FT)</td>
<td>92 F3 families cabbage x broccoli</td>
<td>field</td>
<td>15 surveys at 5–15 days intervals</td>
<td>RFLP</td>
<td>Interval mapping, single-factor ANOVA</td>
<td>MapMaker/ QTL 1.1 Minitab’s GLM procedure</td>
<td>FT</td>
<td>PF</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------------------------</td>
<td>-------</td>
<td>-------------------------------</td>
<td>------</td>
<td>----------------------------------</td>
<td>------------------------------------------</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Annual/biennial (PF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19.7</td>
<td>29.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>192</td>
<td>29.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.9</td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>Resistance to</td>
<td>90 F1 clubroot resistant kale B.</td>
<td>glasshouse</td>
<td>6 weeks 2 years</td>
<td>RAPD</td>
<td>Student’s t-test, Kruskal and Wallis test, One way ANOVA</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasmodiophora</td>
<td>oleracea var. acephala C10 B.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>brassicae Woron</td>
<td>oleracea var. betty41-4.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19.2</td>
<td>29.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.9</td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>Resistance to</td>
<td>91 DH-lines cabbage B. oleracea</td>
<td>phytotron chamber</td>
<td>6 weeks</td>
<td>RFLP</td>
<td>Multiple-QTL Model (MQM), Kruskal and Wallis test</td>
<td>GENSTAT5 MAPQTL 3.0</td>
<td>pb-3</td>
<td>pb-4</td>
<td></td>
</tr>
<tr>
<td>Plasmodiophora</td>
<td>var. capitata “Bindsachsener” x B.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>54</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>brassicae Woron</td>
<td>oleracea var. italicca “ Greenia”</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LG3, 4NE11a; LG1, 2NA8c; LG6, 2NA1b-A20</td>
<td>Voorips et al. (1997)</td>
<td></td>
</tr>
<tr>
<td>Flowe\ing time</td>
<td>149 DH lines B. oleracea var.</td>
<td>field (two trials)</td>
<td>until the first flower</td>
<td>RFLP</td>
<td>Marker regression, ANOVA, SIM function, Interval mapping</td>
<td><a href="http://web.bham.ac.uk/g.g.seaton/MQTL">http://web.bham.ac.uk/g.g.seaton/MQTL</a></td>
<td>FTO2.1 FTO3.1 FTO5.1 FTO5.2 FTO9.1 FTO9.2</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>alboglabra (A12DHd) x B. oleracea</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>var. italicca (GDDH33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Resistance to</td>
<td>138 F1 cabbage B. oleracea var.</td>
<td>infested field</td>
<td>60 days after transplanting</td>
<td>RAPD</td>
<td>RFLP Isozyme morphological</td>
<td>Interval mapping</td>
<td>QGENE</td>
<td>-</td>
<td>30</td>
</tr>
<tr>
<td>Plasmodiophora</td>
<td>capitata x kale B. oleracea var.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LG3, WG6A1- WG1G5 (2.6)</td>
<td>Mori-guchi et al. (1999)</td>
<td></td>
</tr>
<tr>
<td>brassicae Woron</td>
<td>acephala</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pathotype 1 and 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait</td>
<td>Plant Material</td>
<td>Spatial Conditions</td>
<td>Temporal Conditions</td>
<td>Marker Statistical approach</td>
<td>Software</td>
<td>QTL</td>
<td>Variation explained (%)</td>
<td>QTL position LG/Flanking markers/distances (cM)</td>
<td>Authors</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>---------------------</td>
<td>----------------------------</td>
<td>----------</td>
<td>-----</td>
<td>-------------------------</td>
<td>------------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Flowering time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late flowering (LF)</td>
<td>79 substitution lines B. oleracea A12DHd x GDDH33</td>
<td>field (eight trials)</td>
<td>until the first flower</td>
<td>RFLP</td>
<td>ANOVA, Tukey-Kramer multiple comparison tests</td>
<td>-</td>
<td></td>
<td>O1, (0–30.3); O1, (30.3–38.1); O2, (65–80); O3, (0–35); O3, (35–75); O5, (0–34.2); O9, (0.0–43.0); O9, (43.0–64.4); O9, (70.8–106.8); O9, (23.1–47.0)</td>
<td>Rae et al. (1999)</td>
</tr>
<tr>
<td>Early flowering (EF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curd-related traits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Days to budding (dfb)</td>
<td>247 F, rapid cycling B. oleracea (RCB) x B. oleracea var. Cantanese (CAN), 250 F_2 RCB x B. oleracea var. Pusa Katki (PK), 246 F_2 RCB x B. oleracea var. Bugh Kana (BK)</td>
<td>field</td>
<td>180 days after planting on the first flowering day</td>
<td>RFLP</td>
<td>Interval mapping, multiple QTL model</td>
<td>db</td>
<td>20.9 39.6 11.0 1.2 9.9 65.3 6.5 87.6 44.4 44.1 13.9 5.9 21.9 27.9 10.0 39.7 7.8 15.9 9.4 43.8 7.3 4.6</td>
<td></td>
<td>Lan and Paterson (2000)</td>
</tr>
<tr>
<td>First-rank branching (rk1)</td>
<td>Side branching (sidb)</td>
<td>Cluster width (cluw)</td>
<td>Curd width (headw)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------------</td>
<td>---------------------</td>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.6</td>
<td>CAN, C4, EST453g+3;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63.6</td>
<td>PK, C3, EST411c+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76.6</td>
<td>PK, C4, EST373a+16;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.4</td>
<td>PK, C7, EW7E01a+1;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.4</td>
<td>PK, C8, EW8D10b+1;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.6</td>
<td>PK, C9, EST131a+13;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>PK, C5, EW8A11a+2;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.8</td>
<td>BK, C7, WG3D11+11;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.2</td>
<td>CAN, C1, EST566a+2;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.2</td>
<td>CAN, C3, EW8F03a+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.3</td>
<td>CAN, C4, EST453g+3;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.9</td>
<td>CAN, C5, EW8F11c+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.2</td>
<td>CAN, C7, WG3D11+8;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.6</td>
<td>CAN, C8, EW8E09c+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34.8</td>
<td>PK, C4, EST122b+6;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.6</td>
<td>PK, C5, EW8A11a+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.2</td>
<td>PK, C6, WR1D12a+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.9</td>
<td>PK, C7, EW6B07a+3;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.2</td>
<td>PK, C1, EST217a+3;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td>PK, C9, EST311a+14;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45.9</td>
<td>BK, C7, EW8C11a+1;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.6</td>
<td>CAN, C4, EST355b+16;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30.4</td>
<td>CAN, C7, EW8C11a+5;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.1</td>
<td>CAN, C8, EW7C10b+7;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37.8</td>
<td>PK, C4, EST429E+1;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.4</td>
<td>PK, C5, EW8A11a+1;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td>PK, C7, EW8F06b+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.8</td>
<td>PK, C9, EWIG03+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.1</td>
<td>CAN, C1, EW5H03+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.9</td>
<td>CAN, C2, EW9F06b+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.0</td>
<td>CAN, C4, EW9E10+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.9</td>
<td>PK, C1, EST12S+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33.4</td>
<td>PK, C3, EST411c+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.6</td>
<td>PK, C4, EW9B02x+0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5-1 contd.
<table>
<thead>
<tr>
<th>Trait</th>
<th>Plant Material</th>
<th>Spatial Conditions</th>
<th>Temporal Conditions</th>
<th>Marker</th>
<th>Statistical approach</th>
<th>Software</th>
<th>QTL</th>
<th>Variation explained (%)</th>
<th>QTL position LG/Flanking markers / distances (cM)</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apical shoot length</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BK, C4, EW4D04w+7; BK, C7, EW8C11a+4; CAN, C1, EW7C10a+0; CAN, C4, EST453g+4; CAN, C5, EW7E08a+1; CAN, C7, WG3D11+6; CAN, C9, EW1G03a+1; CAN, C3, EST22b+4; CAN, C3, EW8E09e+0; PK, C3, EST411c+0; PK, C4, EST122b+4; PK, C5, EW8A11a+0; PK, C7, EW5A12a+0; PK, C9, EW8E09b+2; BK, C1, EW2E07a+12; BK, C4, EW2B12s+10; BK, C7, EW8C11a+1; BK, C7, EW6C11a+0; CAN, C3, EW9E08a+0; CAN, C4, EST453g+0; CAN, C5, EST19b+7; CAN, C7, EW6B07a+0; CAN, C9, EW1G03a+1; PK, C1, EW8C11x+3; PK, C1, WST88c+0; BK, C1, EW2E07a+15; BK, C4, EW4D04w+23; BK, C7, EW8C11a+2; BK, C7, EW3A04b+23; BK, C7, EW5C12f+0</td>
<td></td>
</tr>
<tr>
<td>First branch length</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>Value</td>
<td>Methodology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------------------------</td>
<td>-------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germination:</td>
<td>105 B. oleracea DH lines A12DHd x GDDH33</td>
<td>- time taken by shoots to reach 1.5 cm and by roots to reach 4 cm; up to 15 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water stress MGT</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heat treatment</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conductivity</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seedling growth:</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seed weight</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root weight</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoot weight</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoot: root weight ratio</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root growth rate</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoot growth rate</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root length</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoot length</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoot: root length ratio</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFLP</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFLP</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>microsatellites</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANOVA</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval mapping</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kruskal-Wallis test regression</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marker regression</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genstat 5</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REML</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MapQTL 3.0</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Java Applet V0.70</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(<a href="http://web.bham.ac.uk/g.g.seaton/qtl_ns/brass.html">http://web.bham.ac.uk/g.g.seaton/qtl_ns/brass.html</a>)</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O1, AC-CACJ01, (118);</td>
<td>13.7</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O3, AC-CAAE02, (84.2);</td>
<td>14.5</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O5, pN23E2, (5.0);</td>
<td>16.4</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O6, p010E2, (7.6);</td>
<td>13.7</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O3, pW102J1, (88.2);</td>
<td>16.6</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O9, p016E1, (71.2);</td>
<td>13.7</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O9, AC-CAGJ05, (11.2);</td>
<td>16.6</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O9, p016E1, (71.2);</td>
<td>13.7</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O4, p0147c1N, (20.3);</td>
<td>16.4</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O9, AC-CAGJ05, (47.3);</td>
<td>13.7</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O6, pR64E2, (11.2);</td>
<td>16.6</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O9, AC-CAGJ05, (47.3);</td>
<td>13.7</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O1, AC-CACE08, (93.9);</td>
<td>16.4</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O4, pW188E2N, (37);</td>
<td>13.7</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O6, AA-CATJ08, (20.2);</td>
<td>16.6</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O3, LEW6e1, (16.3);</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O6, pR64E2, (11.2);</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O7, p36E2, (5.9);</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O7, pN64E2, (37);</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O9, AC-CAGJ05, (47.3);</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O3, pN213J2, (61.7);</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O9, pC14 (cos), (88);</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O7, mBN72AJ1, (67.3)</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5-1 contd...
<table>
<thead>
<tr>
<th>Trait</th>
<th>Plant Material</th>
<th>Spatial Conditions</th>
<th>Temporal Conditions</th>
<th>Marker</th>
<th>Statistical approach</th>
<th>Software</th>
<th>QTL Variation explained (%)</th>
<th>QTL position (LG/Flanking markers/distances (cM))</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flowering time</td>
<td>60 F, B. juncea; 60 F, B. oleracea; 95 F, B. rapa</td>
<td>glasshouse</td>
<td>until the first fully expanded flower</td>
<td>RFLP</td>
<td>Interval mapping, Two-factor analysis of variance</td>
<td>MapMaker/ QTL 1.1 JMP 3.02</td>
<td>14 36</td>
<td>J2, (43); J3, (33); J10, (25); J12, (27); J15, (27); J18, (20); O2, (25); O3, (13); O9, (18); R2, (43); R3, (18); R10, (14)</td>
<td>Axelsson et al. (2001)</td>
</tr>
<tr>
<td>Flowering time</td>
<td>426 BC S1 plants of B. rapa RI line (R500(annual) x Per(biennial)) x R500</td>
<td>growth chamber field</td>
<td>144–160 days after sowing for non-vernalized plants; all vernalized plants were recorded</td>
<td>RFLP</td>
<td>Interval mapping</td>
<td>MapMaker v.2.0 (Haldane function)</td>
<td>VFR2</td>
<td>LG8, Jfg9</td>
<td>Kole et al. (2001)</td>
</tr>
<tr>
<td>Plant size</td>
<td>247 F, rapid cycling B. oleracea (RCB) x B. oleracea var. Cantanese (CAN), 250 F, RCB x B. oleracea var. Pusa Katki (PK), 246 F, RCB x B. oleracea var. Bugh Kana (BK)</td>
<td>field</td>
<td>until the appearance of the first flower</td>
<td>RFLP</td>
<td>Interval mapping, multiple QTL model</td>
<td>SAS, Microsoft Excel, Mapmaker QTL</td>
<td>Laml 15.0 33.0 4.5 3.6 13.7 30.3 4.0 20.0 32.4 9.6 36.7 9.3</td>
<td>BK, C4, EW4D04w+43; BK, C7, EW8C11a+0; BK, C6, EW2C08a+14; BK, C9, EST131a+0; CAN, C4, WG3P04b+4; CAN, C7, EW8C11a+4; CAN, C1, EW6F02+1; BK, C1, EW2E07a+16; BK, C7, EW8C11a+0; CAN, C4, EST453g+2; CAN, C7, EW8C11a+4; CAN, C8, EW809c0</td>
<td>Lan and Paterson (2001)</td>
</tr>
<tr>
<td>Petl</td>
<td>19.5</td>
<td>BK, C7, EW8C11a+2;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stl</td>
<td>12.9</td>
<td>CAN, C4, EW7B02+2;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stw</td>
<td>25.7</td>
<td>CAN, C7, EW8C11a+2;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>non</td>
<td>1.5</td>
<td>CAN, C7, EW3A04b+3;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.3</td>
<td>CAN, C8, WR2F06x+7;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>CAN, C1, EST125+9;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.8</td>
<td>CAN, C1, EW5H03+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.7</td>
<td>CAN, C3, EST165d+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>22.1</td>
<td>CAN, C5, EST195b+8;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.3</td>
<td>PK, C1, EST375x+4;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.3</td>
<td>PK, C3, EW5F09a+6;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.5</td>
<td>PK, C4, EST122b+2;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18.7</td>
<td>BK, C2, EW7B04c+16;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.1</td>
<td>BK, C4, EW2B12s+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>32.2</td>
<td>BK, C7, EW8C11a+2;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.6</td>
<td>BK, C8, EW5G04b+13;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.9</td>
<td>BK, C9, EW6B07d+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.7</td>
<td>BK, C6, EW208a+2;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.7</td>
<td>CAN, C4, EST155b+14;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.0</td>
<td>CAN, C5, EST195b+10;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>33.3</td>
<td>CAN, C7, EW8C11a+2;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>32.5</td>
<td>PK, C4, EST133b+7;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.0</td>
<td>PK, C7, EW5A12a+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.4</td>
<td>PK, C9, EST131A+15;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.3</td>
<td>PK, C5, EST453d+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.7</td>
<td>BK, C7, EW8C11a+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.7</td>
<td>CAN, C3, EW8F03a+3;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16.5</td>
<td>CAN, C7, EW8C11a+4;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.8</td>
<td>CAN, C4, EST155b+6;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.3</td>
<td>CAN, C6, EW5C05c+0;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17.7</td>
<td>PK, C4, EW9E10+4;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.1</td>
<td>PK, C9, EW8E11a+0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait</td>
<td>Plant Material</td>
<td>Spatial Conditions</td>
<td>Temporal Conditions</td>
<td>Marker Statistical approach</td>
<td>Software</td>
<td>QTL Variation explained (%)</td>
<td>QTL position LG/Flanking markers/ distances (cM)</td>
<td>Authors</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>----------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-----------------------------</td>
<td>----------</td>
<td>---------------------------</td>
<td>-----------------------------------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Agrobacterium rhizogenes-mediated transformation efficiency</td>
<td>73 DH lines <em>B. oleracea</em> var. <em>alboglabra</em> (A12DHd) x <em>B. oleracea</em> var. <em>italica</em> (GDDH33)</td>
<td>phytotron chamber</td>
<td>35 days after inoculation</td>
<td>RFLP AFLP microsatellite</td>
<td>Marker regression, interval mapping, ANOVA</td>
<td>QTL CAFE <a href="http://web.bham.ac.uk/g.g.seaton/">http://web.bham.ac.uk/g.g.seaton/</a></td>
<td>8.8 5.4 4.3 29.9 8.3 7.8</td>
<td>GFP-fluorescent root production; LG1, (100); LG3, (85); LG7, (62); Adventitious root production; LG3, (34 and 92); LG5, (38); LG7, (66)</td>
<td>Cogan et al. (2002)</td>
</tr>
<tr>
<td>Resistance to <em>Albugo candida</em></td>
<td>87 RI lines from <em>B. rapa</em> Per (biennial)x R500 (annual)</td>
<td>glasshouse stage</td>
<td>cotyledon stage</td>
<td>RFLP</td>
<td>Interval mapping</td>
<td>Mapmaker 2.0, Mapmaker/QTL 1.1</td>
<td>AC2 resistance AC7 resistance 89.8 1.7 96.3</td>
<td>LG4, wg6c1a-Pub1; LG2, wg2Af11-cs26af; LG4, wg6c1a-Pub1</td>
<td>Kole et al. (2002)</td>
</tr>
<tr>
<td>Flowering time</td>
<td>78 BC3S1 <em>B. rapa</em> segregating for FR1; 100 BC1S1 <em>B. rapa</em> segregating for FR2; 326 F2 plants <em>B. rapa</em> (fr1/fr1, VFR2/VFR2, FR1/FR1,vfr2/vfr2) (Kole et al. 2001)</td>
<td>field growth chamber</td>
<td>days after sowing to the first open flower (DTF) number of leaves on the main axis at flowering (LN)</td>
<td>RFLP SSR</td>
<td>Composite interval mapping (CIM), two-factor analysis of variance with means weighted</td>
<td>JoinMap 3.0, QTL Cartographer, Proc MIXED of SAS</td>
<td>FR1 VFR1 FR2 80.6 14.0 39.0</td>
<td>R2, BrFLC2, (11.4); R2, wg6b10, (53.9); R3, BrFLC5, (24.7)</td>
<td>Schranz et al. (2002)</td>
</tr>
<tr>
<td>Developmental characteristics †</td>
<td>Leaf traits</td>
<td>86 DH lines cauliflower B. oleracea var. botrytis x Brussel sprout B. oleracea var. gemmifera</td>
<td>field</td>
<td>78, 142 and 162 days after sowing</td>
<td>RFLP AFLP SSR</td>
<td>Marker regression, ANOVA</td>
<td><a href="http://web.bham.ac.uk/g.g.seaton/">http://web.bham.ac.uk/g.g.seaton/</a></td>
<td>LW</td>
<td>LG O6, (16); LG O7, (16); LG O8, (28); LG O8, (80); LG O3, (44); LG O6, (16); LG O7, (26); LG O1, (28); LG O1, (70); LG O2, (6); LG O1, (28); LG O1, (0); LG O3, (94); LG O2, (30); LG O6, (4); LG O8, (4); LG O1, (32); LG O5, (28); LG O5, (46); LG O7, (18); LG O8, (24); LG O8, (22); LG O4, (6); LG O7, (56); LG O8, (14); LG O9, (62); LG O1, (22); LG O7, (22); LG O1, (34); LG O1, (68); LG O7, (48); O8, (24); Sebastian et al. (2002)</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------</td>
<td>---------------------------------</td>
<td>------</td>
<td>---------------------------------</td>
<td>--------------</td>
<td>--------------------------</td>
<td>---------------------------------</td>
<td>-----</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Leaf traits</td>
<td></td>
<td>86 DH lines cauliflower B. oleracea var. botrytis x Brussel sprout B. oleracea var. gemmifera</td>
<td>field</td>
<td>78, 142 and 162 days after sowing</td>
<td>RFLP AFLP SSR</td>
<td>Marker regression, ANOVA</td>
<td><a href="http://web.bham.ac.uk/g.g.seaton/">http://web.bham.ac.uk/g.g.seaton/</a></td>
<td>LW</td>
<td>LG O6, (16); LG O7, (16); LG O8, (28); LG O8, (80); LG O3, (44); LG O6, (16); LG O7, (26); LG O1, (28); LG O1, (70); LG O2, (6); LG O1, (28); LG O1, (0); LG O3, (94); LG O2, (30); LG O6, (4); LG O8, (4); LG O1, (32); LG O5, (28); LG O5, (46); LG O7, (18); LG O8, (24); LG O8, (22); LG O4, (6); LG O7, (56); LG O8, (14); LG O9, (62); LG O1, (22); LG O7, (22); LG O1, (34); LG O1, (68); LG O7, (48); O8, (24); Sebastian et al. (2002)</td>
</tr>
<tr>
<td>Flowering traits</td>
<td></td>
<td>86 DH lines cauliflower B. oleracea var. botrytis x Brussel sprout B. oleracea var. gemmifera</td>
<td>field</td>
<td>78, 142 and 162 days after sowing</td>
<td>RFLP AFLP SSR</td>
<td>Marker regression, ANOVA</td>
<td><a href="http://web.bham.ac.uk/g.g.seaton/">http://web.bham.ac.uk/g.g.seaton/</a></td>
<td>LW</td>
<td>LG O6, (16); LG O7, (16); LG O8, (28); LG O8, (80); LG O3, (44); LG O6, (16); LG O7, (26); LG O1, (28); LG O1, (70); LG O2, (6); LG O1, (28); LG O1, (0); LG O3, (94); LG O2, (30); LG O6, (4); LG O8, (4); LG O1, (32); LG O5, (28); LG O5, (46); LG O7, (18); LG O8, (24); LG O8, (22); LG O4, (6); LG O7, (56); LG O8, (14); LG O9, (62); LG O1, (22); LG O7, (22); LG O1, (34); LG O1, (68); LG O7, (48); O8, (24); Sebastian et al. (2002)</td>
</tr>
<tr>
<td>Axillary bud traits</td>
<td></td>
<td>86 DH lines cauliflower B. oleracea var. botrytis x Brussel sprout B. oleracea var. gemmifera</td>
<td>field</td>
<td>78, 142 and 162 days after sowing</td>
<td>RFLP AFLP SSR</td>
<td>Marker regression, ANOVA</td>
<td><a href="http://web.bham.ac.uk/g.g.seaton/">http://web.bham.ac.uk/g.g.seaton/</a></td>
<td>LW</td>
<td>LG O6, (16); LG O7, (16); LG O8, (28); LG O8, (80); LG O3, (44); LG O6, (16); LG O7, (26); LG O1, (28); LG O1, (70); LG O2, (6); LG O1, (28); LG O1, (0); LG O3, (94); LG O2, (30); LG O6, (4); LG O8, (4); LG O1, (32); LG O5, (28); LG O5, (46); LG O7, (18); LG O8, (24); LG O8, (22); LG O4, (6); LG O7, (56); LG O8, (14); LG O9, (62); LG O1, (22); LG O7, (22); LG O1, (34); LG O1, (68); LG O7, (48); O8, (24); Sebastian et al. (2002)</td>
</tr>
<tr>
<td>Stem traits</td>
<td></td>
<td>86 DH lines cauliflower B. oleracea var. botrytis x Brussel sprout B. oleracea var. gemmifera</td>
<td>field</td>
<td>78, 142 and 162 days after sowing</td>
<td>RFLP AFLP SSR</td>
<td>Marker regression, ANOVA</td>
<td><a href="http://web.bham.ac.uk/g.g.seaton/">http://web.bham.ac.uk/g.g.seaton/</a></td>
<td>LW</td>
<td>LG O6, (16); LG O7, (16); LG O8, (28); LG O8, (80); LG O3, (44); LG O6, (16); LG O7, (26); LG O1, (28); LG O1, (70); LG O2, (6); LG O1, (28); LG O1, (0); LG O3, (94); LG O2, (30); LG O6, (4); LG O8, (4); LG O1, (32); LG O5, (28); LG O5, (46); LG O7, (18); LG O8, (24); LG O8, (22); LG O4, (6); LG O7, (56); LG O8, (14); LG O9, (62); LG O1, (22); LG O7, (22); LG O1, (34); LG O1, (68); LG O7, (48); O8, (24); Sebastian et al. (2002)</td>
</tr>
<tr>
<td>Trait</td>
<td>Plant Material</td>
<td>Spatial Conditions</td>
<td>Temporal Conditions</td>
<td>Marker</td>
<td>Statistical approach</td>
<td>Software</td>
<td>QTL Variation explained (%)</td>
<td>QTL position LG/Flanking markers/ distances (cM)</td>
<td>Authors</td>
</tr>
<tr>
<td>-------</td>
<td>----------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>--------</td>
<td>----------------------</td>
<td>----------</td>
<td>-----------------------------</td>
<td>-----------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>Agrobacterium-mediated transformation</strong></td>
<td>Selected lines from AG DH population of <em>B. oleracea</em>, 16 SLs GD33/ A12 for O1, O3 and O7 groups</td>
<td>in vitro culture</td>
<td>5-week-old plants</td>
<td>RFLP</td>
<td>Generalized linear model with Poisson error and log-link function, Analysis of variance</td>
<td>-</td>
<td>Transgenic-root production, Adventitious- and transgenic-root production</td>
<td>O1, (99.3–108.6); O3, (69.3–81.2)</td>
<td>Cogan et al. (2004)</td>
</tr>
<tr>
<td><strong>Regeneration from protoplasts</strong></td>
<td>128 F₂ broccoli <em>B. oleracea</em> var. <em>italica</em> × rapid-cycling <em>B. oleracea</em> (RC)</td>
<td>phytotron chamber</td>
<td>5, 10 and 15 weeks after transfer of microcalli</td>
<td>AFLP microsatellite</td>
<td>Interval mapping, modified bulked segregant analysis</td>
<td>JoinMap 2.0, PROC GLM SAS, PROC CORR SAS, PlabQTL</td>
<td>19.3% 28.6% 28.8% 10 weeks 15 weeks</td>
<td>O₂⁻/C8, 16 cM; O₉/C7, 11 cM; O₂/C8, 17 cM; O₉/C7, 11 cM; O₂/C8, 17 cM</td>
<td>Holme et al. (2004)</td>
</tr>
<tr>
<td><strong>Resistance to Plasmopara brassicae</strong></td>
<td>Ms6 K92 Pb137–522 K92–16 eH 266 F₂, resistant Kale <em>B. oleracea</em> var. <em>acephala</em> C10 × susceptible broccoli <em>B. oleracea</em> var. <em>italica</em> HDEM</td>
<td>glasshouse</td>
<td>6–8 weeks</td>
<td>RAPD RFLP ACGM (amplified consensus genetic markers)</td>
<td>Interval mapping (IM), composite interval mapping (CIM), ANOVA, Pearson coefficient</td>
<td>PROC GLM, QTL Cartographer</td>
<td>Pb-Bo1 6.8 3.0 2.1 6.8 4.3 3.0 2.1 68 2.8 3.3 21.0 4.2 6.1 5.5 34.0 4.3 77.4 2.5 10.7 4.7 68 2.8 3.3 21.0 4.2 6.1 5.5</td>
<td>LG1, T2, (20.41); LG2, S07.1900, (85.41); LG5, PBB7b, (0.01); LG5, ab1.1350, (116.81); LG9, a04.1900, (44.78); LG1, T2, (20.41); LG2, PBB38a, (66.01); LG8, c01.980, (36.49); LG1, T2, (20.41); LG3, a07.1400, (144.46); LG1, T2, (20.41); LG4, a9.983, (112.89); LG5, s07.200, (29.27); LG1, a05.800, (0.01); LG2, r10.1200, (88.21); LG5, PBB7b, (0.01); LG5, a18.1400, (118.19); LG9, aj16.570, (10.01)</td>
<td>Roche-rieux et al. (2004)</td>
</tr>
<tr>
<td><strong>Agrobacterium tumefaciens susceptibility</strong></td>
<td>59 plants of AGDH population <em>B. oleracea</em></td>
<td>in vitro culture</td>
<td>5-days-old seedlings were inoculated and after 50 days evaluated</td>
<td>RFLP</td>
<td>Generalized linear model with logit link and binomial error</td>
<td>JoinMap 3.0, MapQTL 4.0, Genstat 5.0</td>
<td>-</td>
<td>O9, pW233, (40.1)</td>
<td>Sparrow et al. (2004a)</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>-------------------------------------------</td>
<td>-----------------</td>
<td>------------------------------------------------------------</td>
<td>-------</td>
<td>-----------------------------------------------------------</td>
<td>-------------------------------</td>
<td>---</td>
<td>----------------</td>
<td>-------------------</td>
</tr>
<tr>
<td><strong>Transgenic and adventitious root production after <em>Agrobacterium rhizogenes</em> treatment</strong></td>
<td>48 DH plants of NG population <em>B. oleracea</em> var. <em>botrytis</em> (N) x <em>B. oleracea</em> var. <em>geminifera</em> (G)</td>
<td>in vitro culture</td>
<td>6-days-old seedlings were inoculated and evaluated after 35 days</td>
<td>Marker regression, Interval mapping, ANOVA</td>
<td>QTL CAFÉ 1 QTL, 15.7–37.5</td>
<td></td>
<td>15.9</td>
<td>LG1, (34–38) 1.23; LG8, (2–8) 1.23; LG9, (0–4) 1.23; LG9L, (58–88) 1.23</td>
<td>Oldacres et al. (2005)</td>
</tr>
<tr>
<td><strong>Resistance to <em>Plasmodiophora brassicae</em></strong></td>
<td>94 F1,2,3 lines of <em>B. rapa</em> G004 DH line x nou7 Chinese cabbage</td>
<td>greenhouse</td>
<td>6 weeks from sowing</td>
<td>SSR RFLP RAPD SNP Indel markers</td>
<td>Interval mapping Mapmaker/exp 3.0, Mapmaker/QTL 1.1</td>
<td>'Wakayama-01' resistance 26.8</td>
<td>18.3</td>
<td>LG7, BRMS-297-BRMS-088, (~40); LG6, BRMS-100-BRMS-096, (~60); LG2, BN288D-WE24-1, (~10); LG7, BRMS-297-BRMS-088, (~40); LG2, BN288D-WE24-1, (~10)</td>
<td>Suwabe et al. (2006)</td>
</tr>
</tbody>
</table>

Table 5-1 contd...
### Table 5-1 contd...

<table>
<thead>
<tr>
<th>Trait</th>
<th>Plant Material</th>
<th>Spatial Conditions</th>
<th>Temporal Conditions</th>
<th>Marker</th>
<th>Statistical approach</th>
<th>Software</th>
<th>QTL Variation explained (%)</th>
<th>QTL position LG/Flanking markers/distances (cM)</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flowering time (FL)</td>
<td>178 F₂, lines <em>B. rapa</em> rapid cycling line RC-144 x Chinese cabbage line CC-156; 71 DH lines Pakchoi PC-175 x Yellow sarson YS-143; 64 DH lines vegetable turnip VT-115 x YS-143; 136 backcross population lines F. plant (VT-115xYS-143) xVT-115</td>
<td>F₂ non-heated greenhouse</td>
<td>harvested siliques</td>
<td>harvested siliques</td>
<td>Interval mapping (IM), Multiple-QTL model mapping (MQM)</td>
<td>MAPQTL 5.0  Mapchart</td>
<td>13.4 8.9–59.3 19.7 9.3 17.9 11</td>
<td></td>
<td>Lou et al. (2007)</td>
</tr>
<tr>
<td>Seed-related trait:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seed pod length (SPL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seed pod width (SPW)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beak length (SBL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seed color (SC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth-related trait:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plant height (PH)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Branches (PB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaf trait:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaf edge shape (LES)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaf trichomes (LT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaf number (LN)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaf lobes (LB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lamina length (LL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lamina width (LW)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Petiole length (PL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaf area (LA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaf index (LI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turnip trait:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turnip formation (TF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turnip shoots (TS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turnip length (TL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turnip width (TWi)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turnip weight (TWe)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Flowering time

<table>
<thead>
<tr>
<th>Variety</th>
<th>Days to flowering (DTF)</th>
<th>RFLP/SSR/SCAP/CAPS</th>
<th>Composite interval mapping (CIM)</th>
<th>JoinMap 3.0, MapQTL 2.0, QTL Cartographer 1.16</th>
<th>BoFLC2</th>
<th>RFLP/SSR/SCAP/CAPS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B. oleracea var. italica x B. oleracea var. capitata</strong></td>
<td>134 F, 2</td>
<td>RFLP</td>
<td>JoinMap 3.0, MapQTL 2.0, QTL Cartographer 1.16</td>
<td>BoFLC2</td>
<td>36.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRAP</td>
<td>JoinMap 3.0, MapQTL 2.0, QTL Cartographer 1.16</td>
<td>MAPCHART, MapQTL 4.0, PROC CORR of SAS</td>
<td>8.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AFLP</td>
<td>JoinMap 3.0, MapQTL 2.0, QTL Cartographer 1.16</td>
<td></td>
<td>4.3</td>
<td></td>
</tr>
</tbody>
</table>

### Resistance to *Xanthomonas campestris pv. campestris*

<table>
<thead>
<tr>
<th>Variety</th>
<th>Days to flowering (DTF)</th>
<th>RFLP/SSR/SCAP/CAPS</th>
<th>Composite interval mapping (CIM)</th>
<th>JoinMap 3.0, MapQTL 2.0, QTL Cartographer 1.16</th>
<th>BoFLC2</th>
<th>RFLP/SSR/SCAP/CAPS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B. rapa R-0-18 x B162</strong></td>
<td>114 F, 2</td>
<td>AFLP</td>
<td>JoinMap 3.0, MapQTL 2.0, QTL Cartographer 1.16</td>
<td>BoFLC2</td>
<td>63.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interval</td>
<td>JoinMap 3.0, MapQTL 2.0, QTL Cartographer 1.16</td>
<td></td>
<td>51.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>mapping, Multiple-QTL mapping, (MQM) Correlation coefficients</td>
<td></td>
<td></td>
<td>58.0</td>
<td></td>
</tr>
</tbody>
</table>

*Okazaki et al. (2007)*
<table>
<thead>
<tr>
<th>Trait</th>
<th>Plant Material</th>
<th>Spatial Conditions</th>
<th>Temporal Conditions</th>
<th>Marker</th>
<th>Statistical approach</th>
<th>Software</th>
<th>QTL</th>
<th>Variation explained (%)</th>
<th>QTL position LG/Flanking markers/distances (cM)</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leafy trait:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Zhao et al. (2007)</td>
</tr>
<tr>
<td>Leaf edge shape (LES)</td>
<td>160 <em>B. rapa</em> accessions</td>
<td>greenhouse</td>
<td>days to flowering for vernalized plants, days to flowering up to 120 for vernalized plants, 5-week-old plants for LPHY and LPHO</td>
<td>AFLP</td>
<td>One-way analysis of variance (ANOVA), correlation analysis, association mapping: Bayesian approach, Markov Chain, Monte Carlo analyses, Fisher’s exact test</td>
<td>GenStat 8.1, structure 2.1, SAS® software</td>
<td>VDF</td>
<td>0.035</td>
<td>R02, pAG/mCAC0154.7, (63.4); R02, pTA/mCAT0230.2, (84.5); R03, pAT/mCCA0135.0, (15.8); R05, pPG/mCA0398.8, (50.1); R05, pPG/mCA0398, (50.1); R07, pAG/mCAC316.7, (34); R08, pAG/mCAC090.5, (96.5)</td>
<td></td>
</tr>
<tr>
<td>Leaf number (LN)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flowering time:</td>
<td>Flowering time without vernalization (NDF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flowering time after vernalization (VDF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phytate:</td>
<td>Phytate level in leaves (LPHY)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphate level in leaves (LPHO)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phytate level in seeds (SPHO)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphate level in seeds (SPHY)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoot calcium and magnesium content</td>
<td>376 <em>B. oleracea</em> accessions; 74 F1 hybrid cultivars 90 DH lines from AGDH population of <em>B. oleracea</em>, 20 SLs from AGSL population,</td>
<td>glasshouse field</td>
<td>Shoots were sampled after 39 and 105 days of growth</td>
<td>RFLP, AFLP</td>
<td>Composite interval, mapping (CIM), marker regression analysis</td>
<td>QTL Cartographer 2.0, QTL CAFE, GenStat</td>
<td>Mg-f, Ca-m-mr, Mg-m-mr, Ca-g, Ca-f</td>
<td>C2, pN63E1-pW141E1, (70–100); C6, pW197E2-pO104E2, (25–47); C6, pO10E1-pO104E2, (20–55); C7, pO87E2-pN86E1, (0–24); C7, pN86E1-pN20e2, (20–33); C7, pN64E2-pCenE3, (45–75); C8, pCenE3-pN64E2, (75–100)</td>
<td>Broadley et al. (2008)</td>
<td></td>
</tr>
</tbody>
</table>
selected AGSLs were backcrossed and selfed to BC$_{2/3}$

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ca-m</td>
<td>Ca-m</td>
<td>Ca-g-mr</td>
<td>Ca-m-mr</td>
<td>Ca-f-mr</td>
<td>Mg-g-mr</td>
</tr>
<tr>
<td></td>
<td>C7, pN97J2-pCene3, (62–73); C8, pW138J1-pR97J1, (20–52); C8, pW188J1-AC-CAAEB05, (32–56); C8, pW188J1-AC-CAAEB05, (34–60); C8, pW188J1- pO143E2, (40–62); C8, pW188J1-AC-CAAEB05, (38–58); C8, pW188J1-pN173E1, (40–74); C8, pW188J1-pO143E2, (38–64); C9, pN52E2-pO125E1N, (0–10); C9, pO160E1-pW200J1, (70–100); C9, pW137J1-pO119J1, (23–42); C9, pO119J1-pW233J1, (42–55); C9, pW233J1-pO160E1, (54–72); C9, pO7E1-pN47E4NM, (73–90); C9, pW114E2-pW233J1, (35–55); C9, pO7E1-pN47E4NM, (73–90); C9, pW114E2-pW233J1, (35–55); C9, pO119J1-pW233J1, (42–57);</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mg-f-mr</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mg-f</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ca-f</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ca-f-mr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ca-f</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ca-g</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ca-g-mr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table S-1 contd...
<table>
<thead>
<tr>
<th>Trait</th>
<th>Plant Material</th>
<th>Spatial Conditions</th>
<th>Temporal Conditions</th>
<th>Marker</th>
<th>Statistical approach</th>
<th>Software</th>
<th>QTL position</th>
<th>Variation explained (%)</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucosinolate</td>
<td>71 lines from DH38 <em>Brassica napus</em> yellow sarson R500 x pak choi HK Naibaicai, 64 lines from DH30 <em>B. rapa</em> yellow sarson R500 x turnip Kairoku Hakata</td>
<td>glasshouse, growth cabinet</td>
<td>leaves of 40- or 50- days-old plants</td>
<td>AFLP SSR</td>
<td>Interval mapping, MQM mapping, ANOVA</td>
<td>JoinMap 4.0, MapChart 2.1, MAPQTL 5.0</td>
<td>40–56 for DH38</td>
<td>12.7 (DH38) 18 (DH30) 16.4 (DH38)</td>
<td>Lou et al. (2008)</td>
</tr>
<tr>
<td>Taproot thickness (TRT)</td>
<td>Taproot length (TRL)</td>
<td>Taproot weight (TRW)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------</td>
<td>----------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>144 F₂ lines</strong></td>
<td><strong>field</strong></td>
<td><strong>55 days after sowing, 89 days after sowing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. rapa ssp. rapifera QSH97-24 x B. rapa ssp. chinensis AJH97-2</td>
<td><strong>AFLP RAPD</strong></td>
<td><strong>Analysis of variance, trait correlation, Shapiro-Wilk test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. rapa ssp. rapifera</td>
<td><strong>SAS software, Mapmaker/QTL 1.1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>A4-Mid</strong></td>
<td><strong>qTRT2</strong></td>
<td><strong>qTRT3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>16.2 (DH38)</strong></td>
<td><strong>11.6</strong></td>
<td><strong>11.9</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>24.7 (DH30)</strong></td>
<td><strong>1.8</strong></td>
<td><strong>27.4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>24.9 (DH30)</strong></td>
<td><strong>qTRT4a</strong></td>
<td><strong>qTRT4b</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>qTRT5</strong></td>
<td><strong>17.8</strong></td>
<td><strong>19.2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>9.2</strong></td>
<td><strong>17.3</strong></td>
<td><strong>qTRL7</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>qTRL3</strong></td>
<td><strong>qTRL9</strong></td>
<td><strong>qTRW1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>qTRW2</strong></td>
<td><strong>15.9</strong></td>
<td><strong>14.5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>12.2</strong></td>
<td><strong>14.0</strong></td>
<td><strong>qTRW4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>24.8</strong></td>
<td><strong>qTRW5a</strong></td>
<td><strong>18.6</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>qTRW5b</strong></td>
<td><strong>17.1</strong></td>
<td><strong>qTRW9</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>15.4</strong></td>
<td><strong>qTRW9</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5-1 contd...
<table>
<thead>
<tr>
<th>Trait</th>
<th>Plant Material</th>
<th>Spatial Conditions</th>
<th>Temporal Conditions</th>
<th>Marker</th>
<th>Statistical approach</th>
<th>Software</th>
<th>QTL Variation explained (%)</th>
<th>QTL position LG/Flanking markers/ distances (cM)</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resistance to <em>Plasmodiophora brassicae</em></td>
<td><em>B. rapa</em> ssp. <em>pekinensis</em> F$<em>{2/3}$ population K10 x Q5; F$</em>{2/3}$ population C9 x 6R</td>
<td>glasshouse</td>
<td>30 days after sowing</td>
<td>RFLP RAPD AFLP SSR STS</td>
<td>Composite interval mapping (CIM)</td>
<td>JoinMap 3.0, QTL Cartographer 2.5</td>
<td>M85 resistance, K04 resistance</td>
<td>-</td>
<td>R3, HC688, (37.1); R3, HC186, (39.3); R3, E11M2-03, (42.6); R3, OPC11-25, (46.2); R3, HC688, (37.1); R3, HC186, (39.3); R3, E11M2-03, (42.6); R3, OPC11-25, (46.2); R2, E14M3-02, (22.5); R2, m6R, (25.5); R2, E15M4-06, (27.6)</td>
</tr>
</tbody>
</table>

*named according to their location on linkage groups

†located on Sebastian et al. (2000) linkage group

‡Located on Saal et al. (2001) linkage group

Block 1

Block 2

†Abbreviations used by Sebastian et al. (2002): lamina width (LW), lamina petiole length (LPL), bare petiole length (BPL), auricle petiole length (APL), wing petiole length (WPL), leaf apex shape (LAS), midrib width (MW), lobe number (LN), wing number (WN), leaf type (LT), leaf transition zone (LTZ), vernalisation (VN), axillary buds (AB), node number (NN), leaf number (NL), plant high (PH), stem length (SL);

only associations to markers with known map position were described (Zhao et al. 2007);

significant p value in association between measured traits and AFLP markers using multi-QTL model (Zhao et al. 2007)

adventitious root production in the absence of *Agrobacterium*, Oldacres et al. (2005);

adventitious root production in the presence of *A. rhizogenes*, Oldacres et al. (2005);

transgenic root production, Oldacres et al. (2005);

QTL detected by CIM: Ca-f: shoot Ca in field; Ca-g: shoot Ca in glasshouse, Ca-m: mean of shoot Ca in field and glasshouse Broadley et al. (2008);

QTL detected by multiple marker regression have –mr suffix (e.g. Mg-m-mr) Broadley et al. (2008);

Abbreviations used in QTL description by Lou et al. (2008): Ali (aliphatic), Ind (indole), or Aro (aromatic).
previous reports, evidence for genetic interaction was found between two regions affecting flowering time, on LG 6 and LG 8 (Sachan and Singh 1987; Kennard et al. 1994). Homology relationship between genes related to flowering time in *B. napus* (Ferreira et al. 1995) and *B. rapa* (Teutonico and Osborn 1995) was not detected. Comparative mapping analysis found three segments in *B. nigra* genome to be homologous to the top of *A. thaliana* chromosome 5 (Lagercrantz et al. 1996). This is a region where several well characterized flowering time genes such as *FLOWERING LOCUS C* (*FLC*), *FY* and *CONSTANS* (*CO*) are located. In two of those three segments in *B. nigra* flowering time affecting loci were found in close proximity to mapped homologues of the *A. thaliana* *CO* gene. These findings indicated that *B. nigra* *CO* homologs could be the candidate gene for flowering time QTL. Such events have also been identified in *B. oleracea* genome (Bohuon et al. 1998; Table 5-1). The objective of this work was an analysis of an extensive genetic variation in *Brassica* crop genus, exploiting the synteny between and within *Brassica* genomes and *A. thaliana* in order to ease the mapping and identification of latent candidate loci. For this purpose, a population of 150 DH lines of rapid cycling *B. oleracea* derived from a cross between *B. oleracea* var. *alboflagra* x *B. oleracea* var. *italica* was developed, and subsequently scored twice for flowering time. QTL mapping identified six QTL in the following linkage groups: one each on O2 and O3 and two each on O5 and O9. The noteworthy difference in genetic variation was observed, as it was explained in 58% and 93% for the two trials studied, respectively. The observation of homology of regions on O2, O3 and O9, marked with three out of six QTLs, both with each other and in respect to chromosome regions of *B. nigra* (Lagercrantz et al. 1996), and to a single region on *A. thaliana* chromosome 5, pointed at the possibility that one or more of these genes may be candidates for the QTL found in *Brassica*.

The development of backcross substitution lines using the same cross as reported above, and performing longer field trials enabled Rae et al. (1999; Table 5.1) the localization of novel QTLs on chromosomes O1, O2, O3, O5 and O9 of *B. oleracea*. These findings coincided with the results of Bohuon et al. (1998; Table 5-1), in respect to the location of five of QTLs.

Lan and Paterson (2000; Table 5-1) continued the work on localization of flowering time affecting loci (Osborn et al. 1997; Bohuon et al. 1998) and identified at least six additional *Brassica* QTLs on the tops of *B. oleracea* linkage groups 1, 4 and 7, regions homologous to a region of *A. thaliana* chromosome 5.

The collection of a great amount of data on the localization within *Brassica* genomes, chromosomal segments equivalent to the top of the fifth *A. thaliana* chromosome, to which the QTL for flowering time has been mapped (Lagercrantz et al. 1996; Bohuon et al. 1998), raised two important questions. First, how strongly do duplicated genes in polyploid *Brassica*
genomes influence the variation in flowering time traits? Second, are multiple QTLs for a particular trait often controlled by homologous genes? The purpose of the analysis performed by Axelsson et al. (2001; Table 5-1) was to answer these questions. This work was extended and flowering time related QTLs were mapped in *B. juncea*, *B. oleracea* and *B. rapa*, using rapid-cycling lines of each species as the early flowering parent. This study targeted chromosomal regions of *Brassica* that are homologous to chromosomal regions around the *CO* gene in *A. thaliana*. Using molecular markers they found that each diploid *Brassica* species had three regions homologous to the *CO* region of *A. thaliana* and that the amphidiploid *B. juncea* had six copies. Flowering time QTLs were detected in two of the three replicated segments of each diploid (A, B and C) genome, and in three of six replicated segments in the amphidiploid *B. juncea* (two QTLs in the A-genome and one QTL in the B-genome). These findings provided evidence that at least some of the QTLs in the different species are controlled by orthologous and (or) paralogous genes. It is worth mentioning that the QTLs detected on O3 and O9 of *B. oleracea* genome in this study may be correlative to those previously identified by Bohuon et al. (1998).

In contrast to Axelsson et al. (2001; Table 5-1), who reported that the QTL detected correspond to homologues of *CO*, Kole et al. (2001) and Schranz et al. (2002; Table 5.1) provided strong evidence based on confidence intervals for QTL and map positions for *CO* and *FLC*, that the flowering time loci in *B. rapa* are controlled by *FLC* rather than *CO* homologs. This discrepancy may be due to differences in the parentage of populations used to construct linkage maps: Axelsson et al. (2001; Table 5-1) crossed annual types (no vernalization requirement, controlled by *FLC* in *A. thaliana*) to develop segregating populations, whereas Kole et al. (2001) and Schranz et al. (2002; Table 5-1) used segregating populations derived from a cross between biennial and annual form of *B. rapa* (with vernalization requirement).

Additional information came with a study by Sebastian et al. (2002; Table 5.1), described in the previous section, who localized two QTLs affecting vernalization, on linkage groups O7 and O8, respectively. The latter appears to be equivalent to the QTL identified on linkage group 9 of *B. napus* (Ferreira et al. 1995). Other studies have shown the same region in *B. napus* to have a homology with a region in *B. rapa* that also has a putative QTL controlling vernalization (Teutonico and Osborn 1994; Osborn et al. 1997).

Although multiple QTLs affecting various traits related to flowering time have been localized, it was only the first step toward identification of the nature of genetic variation for those traits. In order to achieve this goal, the analysis of candidate genes, as a following step is necessary (Razi et al. 2008). The candidate gene approach here takes advantage of information from many studies, which have identified flowering time
QTLs in different populations and environmental conditions, and from the detailed knowledge of the key genes involved in controlling flowering time that have mainly used *A. thaliana* as a model system (Michaels and Amasino 1999; Sheldon et al. 1999, 2000). Many genes controlling flowering time have been identified in *A. thaliana* but these are considered as major candidate genes *FRI* and *FLC*, which are a key vernalization requirement and response genes, and *CO*, which plays a key role in a photoperiodic regulation of flowering (for a review, see Razi et al. 2008). Keeping in mind, that due to a whole-genome triplication followed by chromosome rearrangements (Lagercrantz and Lydiate 1996; Lysak et al. 2005; Nelson and Lydiate 2006; Ziółkowski et al. 2006), which led to the formation of *Brassica* ancestors, a single locus in the *A. thaliana* genome is represented by multiple copies in *Brassica* genomes. Hence, in *B. oleracea* four *FLC* copies were isolated (Schranz et al. 2002; Lin et al. 2005; Okazaki et al. 2007). A recent study by Okazaki et al. (2007; Table 5-1) reported that among those *FLC* copies (*BoFLC1*, *BoFLC2*, *BoFLC3*, and *BoFLC5*) only *BoFLC2* affected flowering time through vernalization. In amphidiploid *B. napus* in the first instance only five *FLC* genes were identified (Tadège et al. 2001), but more recently Udall et al. (2005) reported the mapping of eight loci, four in each *A* and *C* genomes.

Three *B. oleracea* *FLC* homologs (*BoFLC1*, *BoFLC3*, and *BoFLC5*) were sequenced and their genomic organization characterized (Razi et al. 2008) to infer their identity as candidate genes. Phylogenetic analysis confirmed that *Brassica* *FLC* copies are *A. thaliana* *FLC* orthologs, and supported the occurrence of *FLC* gene duplications before the divergence of *B. oleracea* and *B. rapa* about 4 Mya (Rana et al. 2004). Mapping of four *BoFLC* loci onto the genetic map (A12/GD33 population) confirmed that three copies (*BoFLC1*, *BoFLC3*, and *BoFLC5*) are located within the confidence interval of their respective flowering time QTLs, while *BoFLC4* is located on linkage group O2, where no flowering time QTL was detected. Although the fragment from GD33 parent was considered as a donor of QTL affecting flowering time in backcross progeny obtained, there was no correlation found between late flowering and segregation of GD alleles for *BoFLC1* and *BoFLC3*. These findings together with results indicating that *BoFLC5* and *BoFLC4* were pseudogenes, strongly contradicted the hypothesis that any of them could be candidate for the flowering time QTL in this population.

In *B. rapa*, several QTLs (*VFR1*, *VFR2*, and *VFR3*; *FR1*, *FR2*, and *FR3*) for flowering time were identified in *F₂* and RIL populations derived from a cross between annual and biennial oil types (Teutonico and Osborn 1995; Osborn et al. 1997). *VFR2* was considered to have a large effect, and was suggested to be homologous to *FLC* of *A. thaliana*. Further studies have confirmed the co-localization of several QTLs and loci for *B. rapa* *FLC* homologs *BrFLC2*, *BrFLC3*, *BrFLC5*, and *BrFLC1*, which were then
assigned to linkage groups R02, R03, R03, and R10, respectively (Kole et al. 2001; Schranz et al. 2002; Kim et al. 2006; Lou et al. 2007). Three QTLs for flowering time traits detected by Lou et al. (2007) (Fig. 5-1; Table 5-1) co-localized with LFY (LEAFY), FT (FLOWERING LOCUS T) and VRN2 (VERNALIZATION2) homologs. Among 11 AFLP markers associated with days to flowering, identified in B. rapa through association mapping, three were localized in R02, R03, and R05 (Zhao et al. 2007; Table 5.1). Studies on effects of FLC regions on flowering time have revealed that in B. rapa FR1 (BrFLC2) and VFR2 (BrFLC1) act in a dosage-dependent manner to control the flowering response and respond to vernalization treatment (Schranz et al. 2002; Table 5-1).

5.2.3 Pathogen Resistance

5.2.3.1 Plasmodiophora brassicae

Clubroot, the most important disease of vegetable Brassica crops, is caused by the obligate, biotrophic soil-borne pathogen Plasmodiophora brassicae. The infection leads to gall formation on the root system, which inhibits nutrient and water transport, delays plant growth and increases the susceptibility to wilting. As the protection of these crops by using cultural practices or chemical treatments has proven problematic, the development of resistant cultivars is the most promising way to control clubroot for all Brassica crops (see Chapter 2).

The reported resistance to clubroot concerns different cruciferous species, including the three most extensively cultivated: B. napus, B. rapa and B. oleracea (Crute et al. 1980). In spite of identification of several sources of resistance in B. oleracea (Crute et al. 1980; Crisp et al. 1989; Monteiro and Williams 1989; Dias et al. 1993; Manzanares-Dauleux et al. 2000), only in a few cases has breeding for resistance been successful. This was mainly due to incomplete resistance to clubroot in B. oleracea, which was also mainly found in improper, from a market point of view, types (Crute et al. 1983). Variable expression of clubroot resistance and lack of exact information on genetic control have together limited progress towards developing highly resistant varieties. Some authors have presented evidence, that in B. oleracea this character is determined by several genes with either recessive (Weisæth 1974; Crute et al. 1983; Voorrips and Visser 1993) or dominant alleles (Laurens and Thomas 1993).

Genes for clubroot resistance were for the first time mapped by Landry et al. (1992; Table 5-1). They have identified two QTLs, which explained 61% of the total variation for resistance to race 2 of P. brassicae. The results indicated, that in the case of a cabbage resistant line (B. oleracea var. capitata), alleles for resistance of two QTLs appeared to be dominant over susceptibility.
Figdore et al. (1993; Table 5.1) have detected three QTLs in an F$_2$ population obtained from a cross between broccoli (B. oleracea var. italica), resistant to race 7 of P. brassicae and cauliflower (B. oleracea var. botrytis) susceptible to P. brassicae. They reported one major dominant QTL that came from the broccoli line, one QTL—from the cauliflower line and the third—expressed in the heterozygous state.

Grandclement and Thomas (1996; Table 5-1) investigated resistance to P. brassicae Woron in an F$_2$ population of a cross between a clubroot-resistant kale (B. oleracea var. acephala) and a susceptible cauliflower (B. oleracea var. botrytis). RAPD markers were used for QTL detection. The disease symptoms were observed twice, at different times after inoculation. Three markers were associated with these two observations and three were specific to only one. QTL analysis demonstrated the existence of at least two genetic mechanisms for resistance to P. brassicae.

A significant improvement in QTL mapping and estimating their effects was achieved by Ramsey et al. (1996). They increased the resolution of QTL identification by using molecular markers that limit regions from chromosome of one donor line to be introgressed into recurrent recipient line through subsequent backcrossing and inbreeding. The construction of a library containing B. oleracea recombinant backcross lines with single fragment in the BC$_2$S$_1$ or BC$_3$S$_1$ generations has been proved possible.

QTL analysis for clubroot resistance was performed in a DH population developed from a cross between DH parents of cabbage (B. oleracea var. capitata) and broccoli (B. oleracea var. italica) by Voorrips et al. (1997). They tested each plant by determining symptom grades, and by measuring the fresh weights of the healthy and affected parts of the root system. The analysis performed with multiple QTL mapping approach indicated that at least three QTLs responsible for clubroot resistance were present; two that were found, namely pb-3 and pb-4, that explained only 68% of the difference between the parents and 60% of the genetic variance among DH line means. Contrasting results were obtained by Moriguchi et al. (1999; Table 5-1), who found a single QTL affecting clubroot resistance, performing field experiments on plants naturally infected.

Rocherieux et al. (2004; Table 5-1) analyzed the genetic basis of resistance to clubroot from C10, an inbred line selected from a French kale (B. oleracea var. acephala) landrace, previously reported as highly resistant to several pathotypes of P. brassicae (Manzanares-Dauleux et al. 2000). This line was crossed to a highly susceptible broccoli (HDEM), and genetic control of resistance to five different P. brassicae isolates (four single spore isolates and one field isolate) was investigated in constructed F$_{2/3}$ population. Eight out of nine QTLs identified were isolate-specific and one conferred resistance to all the isolates tested. The analyses showed that the effect of all QTLs was flexible, and relative to the isolates.
The identification of QTL for clubroot resistance has also been conducted in *B. rapa* (Kuginuki et al. 1997; Suwabe et al. 2002). The genetic origin of clubroot resistance was investigated in an $F_3$ population derived from a cross between resistant and susceptible DH lines of Chinese cabbage (*B. rapa* ssp. *pekinensis*) (Suwabe et al. 2003, 2006; Table 5.1). The comparative analysis of three QTL (Crr1, Crr2, and Crr4) identified in the latter study, suggested the co-localization of two major QTL regions, Crr1 and Crr2 in a region of *A. thaliana* chromosome 4 (Fig. 5-2). These results together with previous analysis, which showed the presence of disease-resistance gene cluster in the above mentioned *A. thaliana* region, indicated their common ancestry. Additionally, in *B. oleracea* also two clubroot resistance related QTLs have been reported (Landry et al. 1992; Voorips et al. 1997; Table 5-1). These similarities, together with the hypothesis that both species evolved from a common ancestor, raised the question whether clubroot resistance in *B. rapa* and *B. oleracea* might share the same evolutionary origin and resistance mechanisms.

Recently, a new locus for clubroot resistance was identified by QTL mapping in an $F_2$ population derived from a cross between C9 Chinese cabbage clubroot-resistant DH line and susceptible DH line, 6R (Sakamoto et al. 2008). Locus CRc was localized on linkage group R2. This study also reported the identification of a locus CRk in an $F_2$ population derived from a cross between another resistant line, K10 and susceptible Q5. This locus was located close to Crr3 in linkage group R3. These two clubroot resistant Chinese cabbage DH lines, together with a third one, T136-8, have been developed using resistant genes from CR European fodder turnips (*B. rapa* ssp. *rapifera*), and evaluated on resistance against two isolates of *P. brassicae*, M85 and K04. Additionally, the novel CR loci have been tagged by newly developed sequence tagged site (STS) markers.

### 5.2.3.2 *Albugo candida*

Another pathogen, which causes the disease known as white rust on cultivated *Brassica* species and several wild crucifers, is a biotrophic fungus, *Albugo candida* (Pers.). The symptoms are: appearance of white, blister-like pustules on the lower surface of leaves and on stems, followed with severe distortions in development of the affected organs, especially flowers and siliqua. The consequence of this disease is severe yield losses in turnip rape (*B. rapa*), mustard (*B. juncea*), and rapeseed (*B. napus*) (Kumari et al. 1970; Petrie 1973; Harper and Pittman 1974; Fan et al. 1983). On leafy types of *B. rapa*, such as pak choi (*B. rapa* ssp. *chinensis* L.) and Chinese cabbages (*B. rapa* ssp. *pekinensis* L.), white rust damage can make all the production unsuitable for markets (Santos et al. 2006). Breeding for resistant cultivars
The genetic basis of resistance to white rust in *B. rapa* was investigated by mapping in RIL population derived from a cross between a biennial turnip rape, resistant to AC2 and AC7 races, and an annual sarson, susceptible to both races of *A. candida* (Kole et al. 2002; Table 5-1). The response to both races of *A. candida* was analyzed in young seedlings, in which cotyledons were inoculated and subsequently scored for interactions phenotypes.

is the most successive way to control the disease, besides crop rotation and pesticide spraying.

The genetic basis of resistance to white rust in *B. rapa* was investigated by mapping in RIL population derived from a cross between a biennial turnip rape, resistant to AC2 and AC7 races, and an annual sarson, susceptible to both races of *A. candida* (Kole et al. 2002; Table 5-1). The response to both races of *A. candida* was analyzed in young seedlings, in which cotyledons were inoculated and subsequently scored for interactions phenotypes.
The test found that a single major gene (or tightly linked genes) controlled resistance to the two races. Major resistance locus for both races co-localized on linkage group 4 with previously identified QTL for resistance to race 2 (Kole et al. 1996; Kole et al. 2002). A second minor QTL for AC2 was mapped on linkage group 2. The comparative analysis of map positions of white rust resistance genes in *B. rapa* and *B. napus*, based on common loci in both maps (Parkin et al. 1995; Osborn et al. 1997), pointed to the possibility of location of additional loci, that have not been localized. The synteny of a region with resistance genes mapped in this study to the top or the bottom of *A. thaliana* chromosome 5 was also suggested.

Resistance to white rust was also evaluated in 43 accessions of *B. rapa* ssp. *chinensis* and 19 accessions of *B. rapa* ssp. *pekinensis* at the cotyledon stage with a Portuguese *A. candida* isolate Ac506 collected from turnip leaves (Santos et al. 2006). The two crosses between pak choi BRA 117 (the most resistant accession) and the rapid cycling *B. rapa* line CrC 1.19 were used to analyze the genetic basis of resistance. As a result, dominant/recessive epistatic interaction between two genes affecting the resistance to Ac506 in pak choi BRA 117 was suggested.

### 5.2.3.3 Xanthomonas campestris pv. campestris

The disease that is considered the most serious to some crucifer crops, decreasing both yield and quality (Williams et al. 1980), is a black rot, caused by bacteria *Xanthomonas campestris* pv. *campestris* (Pammel) Dawson. First contact with plant tissues occurs through xylem and mesophyll. The symptoms are marginal leaf chlorosis (usually V-shaped), necrosis and darkening of leaf veins and vascular tissue within the stem, as an effect of an extracellular polysaccharide production by bacteria that stops normal water flow, followed with wilting and necrosis.

In *B. rapa*, the most severe damage is done to turnip and turnip greens (Vicente 2004). The presence of at least six races of the pathogen (Vicente et al. 2001) makes breeding for black rot resistant cultivars a complex approach. Although, races 1 and 4 are thought as the main factors causing disease in *B. oleracea* crops, the extensive screening of *B. oleracea* accessions led to a conclusion that either resistance to these two races did not exist or was very rare (Taylor et al. 2002), in contrast to common resistance to less important races (2, 3, and 6). Because race specific resistance to races 1 and/or 4 is frequently found in other species of the genus *Brassica* (Taylor et al. 2002), which could then serve as a source of this resistance, the inheritance of resistance of *B. rapa* Chinese cabbage accession B162 to the races 1 and 4 of *X. campestris* pv. *campestris* was studied (Soengas et al. 2007; Table 5-1). QTL mapping resulted in identification of four highly significant QTL for
Molecular Mapping of Complex Traits

resistance to both races of *X. campestris* on A06 of linkage map R-o-18 x B162, with two additional QTL for percentage of successful infection by race 4 on A02 and A09. Another piece of evidence indicating the linkage between the genetic factors for resistance to the two races came from correlation of resistance to races 1 and 4 observed in the F$_2$ population. Although, the QTL localization is an initial step to the proposed transfer of resistance from genome A of *B. rapa* to the genome C of *B. oleracea*, promising results of an attempt to introduce resistance to races 1 and 4 from *B. juncea* into *B. oleracea* (Tonguç and Griffiths 2004) suggest that it can be an alternative way to develop resistant cultivars that help to control the disease.

5.2.4 Transformation and Plant Regeneration

5.2.4.1 Agrobacterium-mediated Transformation

Transformation of *Brassica* species with the use of *Agrobacterium*-based system is a commonly used method for gene introduction (reviewed by Poulsen 1996; Puddephat et al. 1996; Sparrow et al. 2006). Efficiency of transformation is limited by genetic factors dependent either on the plant or bacterial genome.

Cogan et al.’s (2002; Table 5-1) paper is the first report on identification of genetic regions that co-regulate this process, apart from the observation indicating the plant genome as a source of genotypic basis to transformation (Gelvin 2000). Using a *B. oleracea* DH mapping population (AG population), they identified seven QTLs affecting transgenic (GFP-green fluorescent protein inserted into the plant using *Agrobacterium rhizogenes*) and adventitious (non-transgenic) root production. Together, the QTLs account for 26% and 32% of the additive genetic variation in the population, respectively. Two of the QTL regions mapped to the linkage groups O3 and O7, controlled both traits. Both parents were the source of regions of the genome carrying QTLs associated with transgenic root production. QTL analysis was performed with two different methods (marker regression and interval mapping), which detected the same regions, except for the region on linkage group O7 for transgenic root production. This work is significant for two reasons: for the first time, QTL mapping approach was used to identify regions associated with transgenic and adventitious root production and also for the first time identified regions that seemed to co-regulate these two processes.

Further analysis of the above mentioned QTL was carried out using a set of A12/GD33 substitution lines resulted in increased precision of localization (Cogan et al. 2004; Table 5-1). This study confirmed co-localization of QTL regions influencing transgenic and adventitious (non-transgenic) root production.
Oldacres et al. (2005; Table 5-1) tested the hypothesis that performance of transgenic root production is associated with performance of adventitious (non-transgenic) root production in *B. oleracea*. In their investigation a second population derived from a cross between cauliflower line Nedcha (*B. oleracea* var. *botrytis*) and the DH Brussels sprout line Gower (*B. oleracea* var. *gemmifera*) (NG population) was used. The regression analysis between the performance of the two traits and the mapping of a QTL controlling the two traits to the same position (LG1, LG8, and LG9) supported the above mentioned hypothesis. Cogan et al. (2002; Table 5-1) also showed two QTLs (LG3 and LG7) to influence both adventitious and transgenic root production in the AG population, with an additional QTL (on LG1) for transgenic root production alone, map on the same linkage group as reported in Oldacres et al. (2005), but in a different position. Hence, these findings provide independent confirmation for the ability of individual QTLs to control both adventitious and transgenic root production.

The genetic loci controlling susceptibility to *A. tumefaciens* in *B. oleracea* were investigated by Sparrow et al. (2004a; Table 5-1), using the same DH mapping population of *B. oleracea* as reported by Cogan et al. (2002; AG population; Table 5-1), and the associated RFLP map. Heritability of a crown gall-formation, as an effect of infection with *A. tumefaciens*, was studied in 8 x 8 diallel arrangement of parental lines that were selected to include a range of susceptibilities to *A. tumefaciens*. The results suggested that although both additive and dominant gene effects control this process, the former was more important. They identified a suggestive QTL correlated with susceptibility to *A. tumefaciens* on linkage group O9, which was further narrowed to a central region, by using substitution lines of the A12 genome with substituted regions of the GD33 genome. The assignment of linkage groups to chromosomes of *B. oleracea* (Howell et al. 2002) revealed the presence of genes correlated with susceptibility to *A. tumefaciens* on chromosome 3 of cytogenetic map of *B. oleracea*.

The next step was the use of both shoot regeneration and susceptibility to *A. tumefaciens* as phenotypic markers and selection of genotypes that could be transformed with high efficiency from the AG population (Sparrow et al. 2004b).

### 5.2.4.2 Plant Regeneration

The ability to regenerate plants from in vitro cultures is important for exploiting transformation and somatic hybridization techniques. Hansen et al. (1999) investigated the genetic control of *B. oleracea* plant regeneration from protoplast-derived microcalli by analyzing F$_2$ progenies from a cross between two parents with high and low regeneration ability. They found that two or three genes control this trait.
The same plant material and the corresponding evaluation data were used by Holme et al. (2004; Table 5-1) to identify two major QTLs connected with the regeneration ability from microcalli of *B. oleracea* protoplasts, and to locate them to linkage groups O2 and O9, and C8 and C7, on two previously published maps, Sebastian et al. (2000) and Saal et al. (2001), respectively. Multiple regression analysis revealed that these two QTLs, both with additive effects from the high-responding parent, accounted for 83% of the total genetic variation for the trait analyzed. The regions carrying those QTLs shared homology with a region from the top of chromosome 5 of *A. thaliana*, known for localization of QTL for plant regeneration and for flowering time (Bohuon et al. 1998; Table 5-1). But in the material used in this study, association between flowering time and regeneration ability was not found.

### 5.2.5 Nutrient Composition

From the agronomic point of view, not only are the loci affecting traits connected with the development of organs suitable for consumers important, but also the nutrient composition of these edible parts including proteins, vitamin C, secondary metabolites such as glucosinolates, and phosphate and other minerals (see Chapters 1.3, 2.2.4.4, 10.4 and 10.5). These traits could be a subject of plant breeding programs, as long as adequate variation can be found among populations and there is sufficient demand from consumers.

Since consumption of plants with insufficient amount of minerals and vitamins is the cause of dietary deficiencies, there are two ways to overcome this limitation. Both relate to plant nutrient enrichment, through agronomic attempts such as the use of fertilizers, or through breeding crops with increased nutrient content. Glucosinolate content is an example of a trait for which the latter is an attractive strategy. Several studies have provided evidence that increased consumption of cruciferous vegetables could significantly reduce the risk of cancer, particularly of the gastro-intestinal tract (Block et al. 1992; Verhoeven et al. 1996). Physiological experiments performed on mammalian cell cultures and rodents, have demonstrated the effectiveness of 3-methylsulfinylpropyl (3-MSP) and 4-methylsulfinylbutyl (4-MSB) isothiocyanates (ITCs), derived from the corresponding glucosinolates in broccoli, as inducers of phase II detoxification enzymes (Zhang et al. 1992; Faulkner et al. 1998). Glucosinolates also have roles in plant defense, flavor and taste (Blau et al. 1978; Kliebenstein et al. 2002).

Mithen et al. (2003) developed ITC-enriched broccoli lines, through introgression of three small segments from *Brassica villosa* genome, a wild relative of *B. oleracea*, into a standard broccoli genetic background. Besides
localization of particular introgressed QTL containing regions, nothing was known concerning the nature of gene or genes within those QTLs.

Recently, the mapping of QTL for glucosinolate content in *B. rapa* leaves was reported (Lou et al. 2008; Table 5.1). This analysis was performed on two DH populations: DH38, derived from a cross between yellow sarson R500 and the pak choi variety HK Naibaicai; and DH30, from a cross between R500 and Kairyou Hakata, a Japanese vegetable turnip variety. Altogether, 16 loci controlling aliphatic glucosinolate accumulation were found, three loci controlling total indolic glucosinolate concentration and three loci regulating aromatic glucosinolate concentrations. The results obtained with comparative analysis between *A. thaliana* and *B. rapa*, together with candidate gene mapping in *B. rapa*, indicated the possibility of several genes involved in the glucosinolate biosynthesis pathway to explain the detected QTLs.

Vegetable Brassicas can serve as a remarkable source of sufficient minerals to satisfy human dietary requirements. However, over half of the world’s population fails to meet these requirements regarding one or more minerals (Welch and Graham 2004). Among group II elements both in plants and animals the most common are calcium (Ca) and magnesium (Mg). Their concentration in shoots and genetic variation underlying these traits were for the first time extensively analyzed in *B. oleracea* by Broadley et al. (2008; Table 5-1). A set of 376 accessions, including examples of wild relatives from natural population and 74 modern F₁ hybrid cultivars, was included in a constructed diversity base set covering a great amount of species-wide allelic diversity within domesticated *B. oleracea*. Homozygous lines from a segregating *B. oleracea* mapping population (AGDH) was used to investigate the genetic basis of variation in shoot Ca and Mg content. QTLs affecting those traits were located on chromosomes C2, C6, C7, C8, and C9. The QTLs on C9 accounted for 14% to 55% of the total genetic variance. The localization of QTL on C9 was further narrowed to a 4 cM region with the use of recurrent backcross substitution lines, derived from the same parents (AGSL population). This region was shown to be homologous to a region on *A. thaliana* chromosome 5 (Parkin et al. 2005), which could serve as a source of putative candidate genes affecting Ca and Mg transport in plants.

In contrast to glucosinolates or minerals, which are the subject of breeding programs of plant enrichment, there is a group of antinutritional substances (e.g., phytate), for which reduction is desired. The differences in phytate and phosphate content was studied by Zhao et al. (2007; Table 5-1) in a different set of 160 *B. rapa* accessions and association mapping was performed to identify genomic regions affecting these traits. The results of association mapping for phytate and phosphate content were compared with that for other traits, which were a subject of human selection. Based on the analysis of marker-marker associations within subpopulations, the presence
of significant LD up to 80 cM was reported. However, LD decay proportional to genetic distance was not observed, though marker density was relatively low. Three models for analysis of association between markers and quantitative traits (simple \( t \)-test, model correcting for population structure, and multi-QTL procedure) were used in this study. Depending on the model tested, between 27 and 170 marker/trait associations were detected. The analysis of the same traits in a set of mapping populations developed from eight parents (Lou et al. 2007; Fig. 5-1; Table 5-1) identified a number of QTL related to different traits. Together, these studies demonstrated that a combined association and linkage mapping approach is effective in identifying regions of the genome controlling important traits in *B. rapa*.

5.3 Future Prospects

5.3.1 Array-based, High-Throughput DNA Marker Systems for *Brassica* Species

5.3.1.1 Single Nucleotide Polymorphism Arrays

The most common form of polymorphism in any genome is nucleotide substitution, also known as single nucleotide polymorphism (SNP; reviewed by Edwards et al. 2007) (see Chapters 4.2.8 and 11). SNP markers are the markers of choice for whole-genome profiling in well-characterized genomes because of their codominance, simplicity (generally biallelic), ease of automation, high reliability and sheer numerical quantity. There are now several automated platforms whereby many thousands of SNPs can be assayed in parallel array (reviewed by Chagné et al. 2007). Agricultural species with large numbers of validated SNP markers include cattle (Khatkar et al. 2007), rice (Shen et al. 2004) and barley (Rostoks et al. 2006). Currently, there are few validated SNPs and no publicly available SNP array available for vegetable *Brassica* species (see Chapter 11). However, the prospects are good that these deficiencies will be redressed in the near future.

Two approaches for large-scale SNP discovery in *Brassica* have recently been reported. The first of these was reported by Duran et al. (2009) and makes use of the large set of *Brassica* EST accessions listed in the GeneBank (907,656 at the time of writing, March 2010) (see Chapter 11). The second takes advantage of “next generation” sequencing technology for deep-sequencing of transcribed genes from European and Chinese parents of a *B. napus* reference mapping population (Trick et al. 2009). The next step—SNP validation—has yet to be completed but initial promising results suggest that a large proportion of SNPs identified by Trick et al. (2009) appear to be heritable in a small sample of the *B. napus* reference mapping population. Heritability is one aspect of SNP validation; another aspect is
informativeness. For a marker to be informative, it must be polymorphic in the target population (breeding or wild), and for measurement of linkage disequilibrium each allele should be present at a frequency greater than 0.1 (Rostoks et al. 2006).

5.3.1.2 Expression Microarrays

Expression microarrays are available for many species, including those for which there are relatively few validated SNP markers (see Chapters 9 and 11). Microarrays consist of many thousands of cloned DNA molecules or oligonucleotides spotted onto a solid matrix. Gene expression analysis involves taking expressed mRNA from a test subject tissue, synthesizing cDNA and then hybridizing to the spotted microarray. More recently, expression microarrays have been used to identify allelic variation—termed single feature polymorphisms (SFPs)—which can be used to measure gene expression and simultaneously genotype test populations (Borevitz et al. 2007). SFPs are based on a range of sequence polymorphisms including SNPs, insertions/deletions and repeat number variation. Several expression microarrays have been developed for Brassica: 95K Agilent array for Brassica species (http://www.cogenics.com/geneexpression/service/brassica.cfm), 94K CombiMatrix array for Brassica species (http://www.intl-pag.org/16/abstracts/PAG16_P07a_708.html) and 24K NimbleGen array for B. rapa (http://www.brassica-rapa.org/BrEMD/). It remains to be seen how effective these arrays will be in detecting SFPs in Brassica species, which have undergone multiple rounds of polyploidization (Nelson and Lydiate 2006) (see Chapter 8).

5.3.1.3 Diversity Arrays Technology

Another option for large-scale, array-based whole-genome profiling in Brassica is diversity arrays technology (DArT, Jaccoud et al. 2001). The advantage of this method over large-scale SNP assay is that no prior knowledge of nucleotide sequence is needed and is therefore an option for any species for which there is little sequence information available. While the DArT marker system is dominant in nature (i.e. presence or absence of marker allele), the hybridization-based assay is highly quantitative as demonstrated in quantitative bulked segregant analysis applications (Wenzl et al. 2007). The dominant scoring of individual markers is somewhat mitigated by the high density of markers (hundreds to thousands per assay) allowing the use of haplotypes for closely linked markers to determine zygosity status for any region. If required, the sequence of the markers can be easily determined from cloned fragments used to spot the array and this sequence can be used to design locus-specific markers. In
contrast to SNP markers where polyploidy makes marker development and validation difficult, the DArT marker system has demonstrated its efficacy in genotyping applications in hexaploid wheat (White et al. 2008) and in hexaploid oats (Tinker et al. 2009). Since *Brassica* diploid species show clear evidence of ancient polyploidy (Lagercrantz and Lydiate 1996), the DArT marker approach may be easier to implement than SNP markers in *Brassica* species. As part of an industry-supported consortium, a DArT array has recently been developed for oilseed *B. napus* and *B. rapa* and currently comprises 2000 polymorphic markers (Andrzej Kilian, pers. comm.). To our knowledge, the array has not yet been tested on vegetable *Brassica* types but is likely to be suitable for the vegetable forms of *B. oleracea* and *B. rapa*.

### 5.3.2 Gene Expression Analysis of Agronomic Traits

QTL analysis has been used extensively for analyzing the genetic control of agriculturally important traits (see section 2. above). A variation on this approach is expression QTL (eQTL) in which global gene expression is analyzed in a mapping population (for details on global gene expression see Chapter 9). Gene transcript levels, when assayed in a mapping population, can be considered as quantitative traits and their variation used to map eQTL. The highest resolution studies employ whole-genome microarrays in reasonably large mapping populations of species with completely sequenced genomes. For example, West et al. (2007) used > 1,000 Affymetrix ATH1 GeneChips (each containing 22,810 genes) to analyze global expression in 211 RILs of Arabidopsis. In that study, 36,871 eQTL were detected: about one-third of which were *cis*-eQTLs (where gene expression is controlled by the gene itself) and two-thirds of which were *trans*-eQTLs (where gene expression is controlled by another locus or loci).

The incredible complexity of gene interactions observed in this controlled and relatively simple population is an indicator of how challenging the eQTL approach will be to understand genetic control of agronomic traits in diverse populations of more complex crop species genomes. Comprehensive expression microarrays for *Brassica* species are now available but so far they have not been used for eQTL analysis.

One potential application of gene expression microarrays in *Brassica* species is investigating the nature of heterosis (see Chapter 2.3.3). This approach was show-cased by an elegant study by Swanson-Wagner et al. (2006) in maize, the archetypal hybrid crop. In that study, global gene expression analysis revealed all possible modes of gene action in maize hybrids compared to inbred parents including overdominance. The challenge with any such analysis will be to tease out the causes of heterosis versus the effects of heterosis.
5.3.3 Elucidating the Genetic Basis of Crop Domestication and Adaptation

Domestication has in every case involved one or several “population bottlenecks” that resulted in reduced genetic diversity in domesticates compared to their wild progenitors (Doebley et al. 2006). This is a problem for plant breeders because availability of adequate genetic diversity is a prerequisite for generating new and better allele combinations. Crop domestication can also be considered a microcosm of species evolution, as Darwin expounded in “The Origin of Species” (1859). Therefore, understanding the genetic basis of crop domestication and adaptation has both basic and applied significance.

There has been some success in identifying (via linkage or association mapping) major domestication genes such as curd development in cauliflower (Purugganan et al. 2000; Smith and King 2000), naked grains in maize (Wang et al. 2005) and large fruit in tomato (Frary et al. 2000). Ross-Ibarra et al. (2007) support an alternative reverse genetic approach that first identifies “signatures of adaptation” in the genomes of domesticated crops and then seeks to identify phenotypic changes associated with those genomic changes. These signatures of adaptation are manifested as lower genetic diversity in genes that have been subjected to selective pressure compared to adaptively neutral genes that have not. With continually improving genotyping tools and the wealth of phenotypic variation present in vegetable Brassicas, both forward and reverse genetic approaches are likely to uncover many of the key genes governing morphological variation and adaptation to domestication.

5.3.4 Epigenetic Regulation of Crop Development

A rapidly growing area of genetics is epigenetics, which can be defined as the study of heritable changes in gene function that occur without a change in the DNA sequence. For an excellent general review of epigenetics in plants, see Grant-Downton and Dickinson (2005). Epigenetics provides a unifying theme that connects an array of what originally seemed to be disparate exceptions to normal Mendelian principles that have accrued over several decades, originally described as paramutation, somaclonal variation, nucleolar dominance, genomic imprinting and transgene silencing. Epigenetic modification is mediated by what has been referred to as the “Three Pillars of Epigenetics”: (1) cytosine DNA methylation; (2) changes in chromatin structure by histone modification; and (3) post-transcriptional gene silencing by short interfering RNAs (Grant-Downton and Dickinson 2005).
The role of epigenetic modification in crop development and performance is not well understood. Indeed, it appears that epigenetic mechanisms are more complex in plants than in other eukaryotes (Grant-Downton and Dickinson 2005). One well known example of an agriculturally significant trait, which is epigenetically controlled, is vernalization—the requirement of a cold period to initiate floral development in some winter annual and perennial plants. Vernalization in Arabidopsis has been shown to be mediated by changes in histone methylation and acetylation of chromatin at the FLC gene (reviewed by Amasino 2005). It may be that many other agronomic traits also have significant epigenetic components.

There have been a few studies of genome-wide DNA methylation and changes in chromatin structure in Brassica species and these have mainly used the low-resolution approach of methylation-sensitive RFLP marker analysis (e.g., Lukens et al. 2006). Until recently, high-resolution methylation analysis was not feasible except in model species. However, the advent of next generation DNA sequencing technology now makes high-resolution methylation analysis (by bisulphite sequencing, Cokus et al. 2008) feasible even for non-model species and is likely to shed further insight into the role of epigenetic mechanisms in crop development.

5.4 Conclusions

The large number of plant QTL that have been mapped to a high resolution will evolve in the coming years to associations of complex phenotypes with their underlying factors. This will create a framework in which to examine how the elusive biological networks interact to create a phenotype (Paran and Zamir 2003).

References

Axelsson T, Shavorskaya O, Lagercrantz U (2001) Multiple flowering time QTLs within several Brassica species could be the result of duplicated copies of one ancestral gene. Genome 44: 856–864.


250 Genetics, Genomics and Breeding of Vegetable Brassicas


Molecular Mapping of Complex Traits


Molecular Mapping of Complex Traits


Molecular Mapping of Complex Traits


ABSTRACT

Cytogenetics of the crop *Brassica* species will soon celebrate 100 years of history. Although the objectives of cytogenetic studies have changed in the course of time, the field is experiencing a renaissance with the advent of new techniques (multicolor fluorescence in situ hybridization, chromosome painting) and due to expanding genomic resources (large-insert DNA libraries, sequence data). Moreover, cytogenomics of Brassicas benefits to a great extent from a variety of genomic tools, resources, and the outstanding genome annotation developed for *Arabidopsis thaliana*. In this chapter we present a brief account of the most significant achievements in *Brassica* cytogenetics and review the latest findings of the genome evolution in *Brassica* crop species. We summarize fundamental facts of the genome composition, homeologous chromosome pairing, chromosome identification, chromosome characterization, and evolutionary dynamics of dispersed and satellite repeats in diploid and allopolyploid *Brassica* genomes. Furthermore, the recent progress in comparative cytogenetic studies by chromosome painting mapping with bacterial artificial chromosomes (BAC) and their impact on the reconstruction of genome evolution in the genus *Brassica* are discussed.

**Keywords:** cytogenetics, fluorescence in situ hybridization (FISH), comparative chromosome painting (CCP), chromosomal rearrangements, Ancestral Crucifer Karyotype (ACK), hexaploidization, *Brassica* genome evolution

---

1Department of Biotechnology, Adam Mickiewicz University, Poznań, Poland.
2Department of Functional Genomics and Proteomics, Masaryk University, Brno, Czech Republic.
3Department of Plant Breeding and Biotechnology, Swedish University of Agricultural Sciences, Alnarp, Sweden.
*Corresponding author: pzio@amu.edu.pl
6.1 Classical Cytogenetic Approaches and the Genetic Background of the Brassica Genus

Despite relatively tiny and numerous chromosomes, cytogenetic studies on Brassica species have a very long history. The first chromosome analyses in the genus are from the beginning of the 20th century (Takamine 1916; Karpechenko 1922; Morinaga 1928, 1933, 1934; U 1935). Takamine (1916), Karpechenko, (1922) and Morinaga (1933, 1934) reported the chromosome numbers of the economically important species within the Brassica genus. Since that time, chromosome numbers have been determined for more than 1,500 different Brassicaceae taxa (summarized in Warwick and Al-Shehbaz 2006). Chromosome numbers for vegetable Brassicas are shown in Table 6-1.

Meiotic analysis of chromosome pairing and segregation in Brassica interspecific crosses elucidated evolutionary relationships among the genomes of diploid and allotetraploid species within the genus (Morinaga 1933, 1934; U 1935). It was suggested that the natural amphidiploid (an allotetraploid formed by hybridization of two species followed by chromosome doubling; in consequence it appears to be a functional diploid) Brassica species originated from interspecific crosses between diploid species (Fig. 6-1; U 1935). According to this hypothesis, the diploid species Brassica rapa (the so called genome A), and diploid B. nigra (genome B) are parents of natural amphidiploid B. juncea (genomes A and B), B. rapa and B. oleracea (genome C) are parents of the amphidiploid B. napus (genomes A and C), whereas B. nigra and B. oleracea formed the amphidiploid B. carinata (genomes B and C). These relationships are known as the “U-triangle” (U 1935) and have been widely accepted, since they were confirmed by many subsequent genetic and biochemical analyses.

6.2 Addition and Substitution Lines

Access to a natural or synthetic amphidiploid is a prerequisite for the development of mono- and disomic addition lines of diploid species. Crossing the amphidiploid to one of the parental species results in a sesquidiploid hybrid, with two sets of one genome and one set of the other genome. Selfing or recurrent backcrossing generates aneuploids comprising monosomic addition lines. Disomic addition lines arise in the progeny of monosomics. Characterization of the addition lines spans over the documentation of plant morphology, cytology, isozyme patterns, and genome-specific DNA markers that distinguish the various alien chromosomes. Monosomic addition lines offer unique opportunities for cytogenetic analysis (Chang and de Jong 2005). They are well suited for localization of genes and identification of linkage groups, definition
**Table 6-1** Chromosome numbers for vegetable Brassicas (data from Warwick and Al-Shehbaz 2006). Some other important species of the Brassicaceae family are included for comparison.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Common name</th>
<th>Genome structure</th>
<th>Ploidy level and chromosome number</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Leaf Brassicas</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. oleracea var. capitata</em></td>
<td>cabbage</td>
<td>CC</td>
<td>$2n = 2x = 18$</td>
</tr>
<tr>
<td><em>B. oleracea var. gemmifera</em></td>
<td>Brussels sprouts</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. oleracea var. acephala</em></td>
<td>kale, collard greens</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. oleracea var. alboflabra</em></td>
<td>Chinese broccoli (kai-lan)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. rapa subsp. pekinensis</em></td>
<td>Chinese cabbage</td>
<td>AA</td>
<td>$2n=2x=20$</td>
</tr>
<tr>
<td><em>B. rapa subsp. chinensis</em></td>
<td>bok-choy (pak-choi)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. rapa var. pervidis</em></td>
<td>komatsuna</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. rapa subsp. narinosa</em></td>
<td>tatsoi</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. juncea</em></td>
<td>mustard greens</td>
<td>AABB</td>
<td>$2n=4x=36$</td>
</tr>
<tr>
<td><strong>Stem Brassicas</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. oleracea var. gongylodes</em></td>
<td>kohlrabi</td>
<td>CC</td>
<td>$2n=2x=18$</td>
</tr>
<tr>
<td><strong>Flowerhead Brassicas</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. oleracea var. italica</em></td>
<td>calabrese, broccoli</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. oleracea var. botrytis</em></td>
<td>cauliflower, Romanesco broccoli</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Root Brassicas</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. rapa</em></td>
<td>turnip</td>
<td>AA</td>
<td>$2n=2x=20$</td>
</tr>
<tr>
<td><em>B. napus var. napobrassica</em></td>
<td>swede (rutabaga)</td>
<td>AACC</td>
<td>$2n=4x=38$</td>
</tr>
<tr>
<td><strong>Other Brassicas</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. nigra</em></td>
<td>black mustard</td>
<td>BB</td>
<td>$2n=2x=16$</td>
</tr>
<tr>
<td><em>B. napus subsp. napus</em></td>
<td>rapeseed</td>
<td>AACC</td>
<td>$2n=4x=38$</td>
</tr>
<tr>
<td><em>B. carinata</em></td>
<td>Ethiopian mustard</td>
<td>BBCC</td>
<td>$2n=4x=34$</td>
</tr>
<tr>
<td><strong>Other Brassicaceae</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Raphanus sativus</em></td>
<td>radish</td>
<td></td>
<td>$2n=2x=18$</td>
</tr>
<tr>
<td><em>Sinapis alba</em></td>
<td>white mustard</td>
<td></td>
<td>$2n=2x=24$</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>thale cress</td>
<td></td>
<td>$2n=2x=10$</td>
</tr>
<tr>
<td><em>A. lyrata</em></td>
<td>lyrate rockcress</td>
<td></td>
<td>$2n=2x$, $4x=16$, 32</td>
</tr>
<tr>
<td><em>Capsella rubella</em></td>
<td>pink shepherd’s purse</td>
<td></td>
<td>$2n=2x=16$</td>
</tr>
</tbody>
</table>
and physical mapping of chromosome-specific DNA markers, analysis of homeologous pairing and monitoring of intergenomic gene transfer. A nullisomic condition or the substitution of a certain chromosome pair by its homeologues, or by another pair, from a related species, has been also informative regarding gene linkage and the genetic control of certain characters.

Addition and substitution lines, as well as a nullisomic, portraying the dissection of the genomes of the vegetable Brassicas (B. oleracea, B. rapa, B. juncea and B. napus) are presented in Table 6-2. Most exploited and informative are the monosomic addition lines resulting from the dissection of B. oleracea. Thus up to seven or eight of the nine chromosome types of B. oleracea have been incorporated in the background of B. rapa (McGrath and Quiros 1990; McGrath et al. 1990; Heneen and Jørgensen 2001) and Raphanus sativus (Kaneko et al. 1987, 1991). The addition lines were morphologically distinguishable when in the R. sativus background, and to a much less extent

Figure 6-1 Overview of genetic relationships between Brassica species, as described by Nagaharu U in 1935 [picture modified from Mike Jones, Wikipedia (http://en.wikipedia.org/wiki/Triangle_of_U)].
Cytogenetic Studies in Vegetable Brassicas

in the *B. rapa* background, apparently due to the fact that *B. oleracea* and *B. rapa* are more closely related. Substitution of a chromosome pair of *B. rapa* by another pair of *B. oleracea* has been also encountered in these materials (McGrath et al. 1990; Chen et al. 1997a).

In addition to plant morphology, analysis of the *B. rapa*-*B. oleracea* monosomic addition lines highlighted further features (references in Table 6-2). Thus, the molecular characterization of these lines characterized synteny groups and disclosed the prevalence of duplicated gene loci, in agreement with reported gene linkage and physical localization maps, and supporting the hypothesis that these diploid species are actually secondary polyploids. The ability to differentiate between the chromosomes of the two species during diakinesis, based on chromatin condensation patterns, made it possible to accurately define if univalents are alien or background chromosomes, and to monitor homeologous pairing in bivalents and multivalents (Fig. 6-2). That the short and long arms of the alien *B. oleracea* chromosome can pair with arms that belong to different *B. rapa* chromosomes is of relevance as to the relatedness and evolution of these genomes. The use of rDNA probes in combination with fluorescence in situ hybridization (FISH) was helpful in defining the chromosomes with rDNA

<table>
<thead>
<tr>
<th>Background genome</th>
<th>Dissected genome</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Additions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. rapa</em></td>
<td><em>B. oleracea</em></td>
<td>Quiros et al. 1987</td>
</tr>
<tr>
<td></td>
<td></td>
<td>McGrath and Quiros 1990</td>
</tr>
<tr>
<td></td>
<td></td>
<td>McGrath et al. 1990</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hu and Quiros 1991</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lee and Namai 1992, 1994</td>
</tr>
<tr>
<td><em>B. oleracea var. alboglabra</em></td>
<td></td>
<td>Chen et al. 1992,1997a,b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cheng et al. 1994a,b, 1995a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jørgensen et al. 1996</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heneen and Brismar 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Heneen and Jørgensen 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hasterok et al. 2005b</td>
</tr>
<tr>
<td><em>Raphanus sativus</em></td>
<td><em>B. oleracea</em></td>
<td>Fantes and MacKay 1979</td>
</tr>
<tr>
<td><em>B. rapa</em></td>
<td><em>B. juncea</em></td>
<td>Struss et al. 1991, 1992</td>
</tr>
<tr>
<td><strong>Substitutions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. rapa</em></td>
<td><em>B. oleracea</em></td>
<td>McGrath et al. 1990</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chen et al. 1997a, 1997b</td>
</tr>
<tr>
<td><em>B. oleracea</em></td>
<td><em>B. napus</em></td>
<td>Chiang et al. 1980</td>
</tr>
<tr>
<td><em>B. juncea</em></td>
<td><em>B. napus</em></td>
<td>Banga 1988</td>
</tr>
<tr>
<td><strong>Nullisomic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>B. juncea</em></td>
<td><em>B. juncea</em></td>
<td>Cheng et al. 2001</td>
</tr>
</tbody>
</table>

Table 6-2 Dissected and background genomes of chromosome addition and substitution lines, and a nullisomic, involving *B. oleracea*, *B. rapa*, *B. juncea* and *B. napus*.
The occurrence of intra- and intergenomic gene recombination and chromosomal structural changes, like deletions and translocations, has been documented during the development of the addition lines and in their progenies. Obtained double monosomics are useful to reveal intragenomic homeology. It was possible to define which chromosomes, and in some cases to infer which regions on chromosomes, that harbor genes of agronomic importance, like those controlling erucic acid content and seed color. At least three chromosomes carried genes that controlled seed color either maternally or biparentally. The transmission of the alien chromosome in the addition lines occurred at different frequencies, and seed setting was variable, but enough for maintenance of these lines.
The potential of developing new cultivars from disomic, double disomic or multiple disomic addition lines of *B. rapa-B. oleracea* has been pointed out (Lee and Namai 1992, 1994). Further use of addition and substitution lines in plant breeding relate to the inheritance of the boltic trait (Banga 1988; Kaneko et al. 2000). Of the 10 chromosome types of *B. rapa*, eight were represented in a *R. sativus* background, and similar to the situation with *B. oleracea*, they gave rise to morphologically differentiable addition lines (Kaneko et al. 1991, 2003). A comparison of a double monosomic *R. sativus* line, harboring one chromosome from each of *B. oleracea* and *B. rapa*, with its parental monosomic addition lines elucidated chromosome homeological relationships among the three species (Kaneko et al. 2002). The dissection of *B. napus* and *B. juncea* in addition and substitution lines, and in a nullisomic, have been restricted to a limited number of chromosome types (Chiang et al. 1980; Banga 1988; Struss et al. 1991, 1992; Cheng et al. 2001).

In summary, chromosome addition and substitution lines and nullisomics are valuable tools for the elucidation of genome composition and the assignment of genes, DNA markers and synteny groups to defined chromosomes. They are also suitable for chromosome morphological and structural studies, physical mapping of genes and specific DNA sequences, monitoring of phylogenetical relationships and intergenomic gene transfer, as well as providing bridging materials for plant breeding.

**6.3 Interspecific Hybridization and Homeologous Chromosome Pairing**

A large number of interspecific and intergeneric sexual and somatic hybrids, involving the Brassicas, has been obtained (see reviews by Warwick et al. 2000; FitzJohn et al. 2007; Prakash et al. 2009). In addition to the goal of expanding the genetic variability for breeding purposes, hybridization has contributed to the elucidation of interrelationships between genomes, expressed by homeologous chromosome pairing (Snowdon 2007). The high crossability among *Brassica* species and close relatives, and the occurrence of homeologous pairing and genetic recombination in many of these hybrids is a suitable means of transferring desirable characters from one species to the other. Only hybrids denoting intergenomic chromosome homeology that pertain to the vegetable Brassicas (*B. oleracea*, *B. rapa*, *B. napus*, *B. juncea* and *B. carinata*), will be considered here.

Intergeneric relationships inferred from meiotic studies are usually based on pairing formulas expressing numbers of univalents, bivalents and multivalents, reflecting the affinities between chromosomes. Most of the reported meiotic pairing formulae are not based on actual chromosome identification of parental genomes, due to lack of differentiating morphological features, or not applying discriminating techniques, like
Genomic in situ hybridization (GISH) or FISH and chromosome painting using chromosome-specific DNA probes. Thus, caution is needed when inferring auto- and allo syndesis of paired chromosomes, and parental origin of univalents, if chromosome identification has not been ascertained. The frequent occurrence of duplications in the *Brassica* genomes contributes to the often documented intragenomic chromosome pairing (Li et al. 2005; Leflon et al. 2006; Ge and Li 2007).

Differences in chromatin condensation patterns of parental chromosomes are useful for the detection of homeologous pairing in bivalents and multivalents, and determination of the parental origin of univalents (Heneen et al. 1995). Discrimination between genomes in hybrids and allopolyploids, intergenomic recombination and introgression have been more accurately determined by the application of FISH and GISH (Li et al. 2005; Wang et al. 2005; Leflon et al. 2006; Ge and Li 2007; Snowdon 2007; Nicolas et al. 2008).

The earliest *Brassica* hybridizations were among the three diploids *B. rapa*, *B. oleracea* and *B. nigra* and their three amphidiploids *B. napus*, *B. juncea* and *B. carinata*. All three diploid basic genomes have been further brought together in trigenomic allohexaploids. Most information on the relationship between homeologous chromosomes has been inferred from hybrids, synthetic allopolyploids, digenomic triploids, and monosomic addition lines. While multivalents occurred in synthetic amphidiploids, diploid-like meiosis achieved during evolution, prevailed in natural polyploids. A genetic control of the diploid-like meiosis in natural *Brassica* allopolyploids has been advocated (Jenczewski et al. 2003 and references therein). Apparently, crossover formation and frequency are not only determined by the level of intra- and intergenomic relatedness, and genetic control, but also by the composition of the chromosome complement, whether constituted of different genomes or a mixture of similar and different genomes (Nagpal et al. 1996; Nicolas et al. 2008).

Analysis of hybrids, synthetic allopolyploids, backcrosses and addition lines involving the six *Brassica* species of the U-triangle led to the conclusion that the A and C genomes are phylogenetically closer to each other than to the B genome. This has been corroborated by molecular DNA analysis (Truco et al. 1996 and references therein). In addition to the established chromosome homeologies between the A, B and C genomes documented in the six cultivated species, homeologies were inferred between the vegetable Brassicas and *B. balearica*, *B. barrelieri*, *B. bourgeaut*, *B. cossoneana*, *B. cretica*, *B. fruticulosa*, *B. gravinae*, *B. incana*, *B. insularis*, *B. macrocarpa*, *B. maurorum*, *B. montana*, *B. oxyrrhina*, *B. rupesris*, *B. tournefortii* and *B. villosa* (Inomata 1993; Bothmer et al. 1995; Prakash et al. 2009). Intergeneric hybrids, exhibiting homeologous chromosome pairing, have been obtained between vegetable Brassicas and various other genera including *Coicya*, *Diplotaxis*, *
Enarthrocarpus, Eruca, Erucastrum, Hirschfeldia, Moricandia, Orychophragmus, Raphanus, Sinapidendron and Synapis (Du et al. 2008; Prakash et al. 2009). In hybrids involving Orychophragmus violaceus, genome separation and elimination of the O. violaceus genome prevailed (Hua et al. 2006; Li and Ge 2007).

Hybridization and backcrossing have been employed for the introgression of desirable characters, such as disease resistance and characters relating to seed quality, into Brassica genomes. Monitoring the incidence of hybridization between crop plants and related wild species, homeologous chromosome pairing and consequent gene transfer, are pertinent issues before marketing of transgenic crops.

6.4 Genome Organization at Chromosome Level—Euchromatin and Heterochromatin

One of the molecular mechanisms modulating the gene function in higher eukaryotes is the make-up of chromatin in their immediate vicinity. At the same time, it is also clear that genes constitute only a fraction of genomes, and that usually a large part of the DNA has a noncoding nature. Chromatin incorporates several classes of protein complexes, which are responsible for a level of its compaction. From a cytogenetic point of view, there are two types of chromatin which can be easily identified under a microscope: euchromatin, a less condensed fraction in which most of transcriptionally active genes are found, and heterochromatin, a fraction that remains highly condensed throughout the cell cycle, occupied mainly by repetitive sequences (reviewed by Richards and Elgin 2002).

In higher eukaryotes heterochromatin is located at telomeres, centromeres and pericentromeric regions; in plants, in addition to these regions, heterochromatin is located at the nucleolar organizer regions (NORs), at knobs and in some cases along B chromosomes (Alfenito and Birchler 1993; Copenhaver et al. 1999; Fransz et al. 2000; McCombie et al. 2000).

Under phase-contrast microscopy, heterochromatin in interphase and prophase nuclei may be clearly recognized as dark regions in a light background of euchromatin; conversely, DAPI (4’,6-diamidino-2-phenylindole) staining reveals bright heterochromatin foci under an epifluorescence microscope. In Brassica species, heterochromatin and euchromatin can be easily distinguished throughout the meiotic prophase I, and in interphase nuclei (see Fig. 6-3). At the mitotic metaphase, chromosomes are usually too compact to discriminate both types of chromatin.

Plants often possess much larger genomes than animals (Bennett and Leitch 2005), mainly due to a higher frequency of the transposable elements.
(TEs) amplification. In consequence, large regions of non-genic DNA are formed, which are subsequently methylated to stop further expansion of TEs (Bennetzen 2005). At the same time, chromatin remodeling converts initially relaxed chromatin of the chromosome region into the compact state, mainly by histone methylation (Lippman et al. 2004; Fransz et al. 2006). This can cause formation of heterochromatin knobs, which in an extreme situation can be the major constituent of all chromosomes in the genome, as in the case of onion Allium cepa (de Jong et al. 1999). However, Brassica genomes are relatively small and compact as compared to many other plant genomes, with chromosome architecture resembling that of Arabidopsis thaliana. Thus, heterochromatin blocks can be found preferentially around centromeres, at NORs, and in a few distal and/or interstitial sites (usually described as knobs or chromomeres). In B. oleracea, only one non-centromeric knob was found, which was further identified as the major 5S ribosomal DNA (rDNA) locus (Ziolkowski and Sadowski 2002). In B. rapa five non-centromeric heterochromatic knobs were reported, one of them bearing a 5S rDNA locus (Koo et al. 2004). Up to now, such an analysis has not been performed in other Brassica species.

6.5 Towards Identification of Brassica Chromosomes

One of the main goals of cytogenetic analysis is to identify and efficiently distinguish chromosomes within the genome(s) under study. The reasons to do so are numerous: at the cytological level, it enables the characterization of chromosomal aberrations, and to assign chromosomes to parental species in
allopolyploids, and introgression lines. This is of major importance in plant breeding strategies, when the identification of genetic material of donor species may significantly accelerate production of new cultivars. On the other hand, identification of alien chromosomes in addition or substitution lines can be beneficial for mapping of quantitative trait loci (QTL), and for interspecies gene transfer in breeding programs. On the molecular level, identification of chromosomes is needed for integration of genetic maps with physical maps produced by karyotype analysis and cytogenetic determination of the position of DNA sequences on chromosomes. This is an important step in advanced genome sequencing programs, where particular bacterial artificial chromosome (BAC) or shotgun-derived contig could be assigned to a chromosome.

Chromosome identification is relatively simple for species with a low number of large chromosomes, which have different morphology and different condensation patterns/distribution of repetitive sequences. In such a case, constructing a karyotype is then possible through classical cytogenetic analyses. Unfortunately, Brassica species possess relatively small genome sizes (~ 650 Mb for B. oleracea; for more details see Table 6-3), and thus their chromosomes are very small. As they measure only 2–5 µm at the mitotic metaphase stage (Cheng et al. 1995a), karyotype construction is a hard task. Because of these difficulties, researchers have been looking for some other possibilities to analyze Brassica chromosomes. One of them is to use pachytene stage of meiosis, where incomplete condensed bivalents are much longer than their counterparts at the mitotic metaphase. The difference in chromosome length between mitotic metaphase and meiotic pachytene chromosomes is 10- to 40-fold, depending on a species (de Jong et al. 1999). For Brassica, the length of pachytene bivalents is 15- to 20-fold the length of mitotic metaphase chromosomes (Ziolkowski and Sadowski 2002). Thus, the resolution can be greatly improved by analysis of pachytene chromosomes.

The first trial to construct a Brassica karyotype dates to 1960. Röbbelen (1960) used pachytene bivalents for constructing karyotypes of three diploid Brassica species on the basis of structural characteristics of total chromosome length, symmetry of the arms, and the shape of the heterochromatic pericentromeric regions. Based on their absolute length, chromosomes were classified into five different types: very short (up to 20 µm), short (20–25 µm), medium (25–30 µm), long (30–40 µm), and very long (more than 40 µm). Based on the centromere position, they were classified as chromosomes with a median or submedian centromere, respectively, and a chromosome with subterminal centromere and a satellite. The work reports the only complete pachytene karyotype for B. oleracea and B. nigra published to date. Recently, a high-resolution karyotype of B. rapa subsp. pekinensis based on
Chromosome banding techniques have been applied to Brassica species starting from the late 1980s (Wang and Luo 1987; Wang et al. 1989). Both C-banding and G-banding techniques revealed that Brassica chromosomes are not enough structurally differentiated to be fully distinguished based on these methods (Nishibayasahi 1992). The main problem was still the length of mitotic metaphase chromosomes. One possible solution was to use late prophase preparations for karyotyping. Olin-Fatih and Heneen (1992) described karyotypes of B. oleracea, B. rapa and B. napus studied using C-banding. The chromosomes were grouped as with a median, submedian, subterminal, and terminal centromere, respectively. According to the authors, all chromosome pairs were morphologically distinguishable. This multicolor FISH using pachytene and mitotic metaphase chromosomes has been published (Koo et al. 2004).

Table 6-3 Genome sizes in Brassica species of the U triangle (Table redrawn from http://www.brassica.info/info/reference/genome-sizes.php with permission).

<table>
<thead>
<tr>
<th>Species (genome)</th>
<th>Genome size (Mb)</th>
<th>Variety/ common name</th>
<th>Method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. rapa (A)</td>
<td>468–516</td>
<td>var. chinensis Pak choi</td>
<td>FCM</td>
<td>Arumuganathan and Earle 1991</td>
</tr>
<tr>
<td></td>
<td>507</td>
<td></td>
<td></td>
<td>Arumuganathan and Earle 1991</td>
</tr>
<tr>
<td></td>
<td>511</td>
<td>var. rapifera turnip</td>
<td></td>
<td>Arumuganathan and Earle 1991</td>
</tr>
<tr>
<td></td>
<td>529</td>
<td></td>
<td></td>
<td>Johnston et al. 2005</td>
</tr>
<tr>
<td></td>
<td>564</td>
<td></td>
<td></td>
<td>Bennet and Smith 1991</td>
</tr>
<tr>
<td></td>
<td>784</td>
<td></td>
<td></td>
<td>Bennet et al. 1982</td>
</tr>
<tr>
<td>B. nigra (B)</td>
<td>468</td>
<td>black mustard</td>
<td>FCM</td>
<td>Arumuganathan and Earle 1991</td>
</tr>
<tr>
<td></td>
<td>632</td>
<td></td>
<td>FCM</td>
<td>Arumuganathan and Earle 1991</td>
</tr>
<tr>
<td></td>
<td>760</td>
<td></td>
<td>FD</td>
<td>Bennet and Smith 1976</td>
</tr>
<tr>
<td>B. oleracea (C)</td>
<td>599–618</td>
<td>var. italica broccoli</td>
<td>FCM</td>
<td>Arumuganathan and Earle 1991</td>
</tr>
<tr>
<td></td>
<td>603</td>
<td>var. capitata cabbage</td>
<td>FCM</td>
<td>Arumuganathan and Earle 1991</td>
</tr>
<tr>
<td></td>
<td>628</td>
<td>var. gemmifera Brussels sprout</td>
<td>FCM</td>
<td>Arumuganathan and Earle 1991</td>
</tr>
<tr>
<td></td>
<td>628–662</td>
<td>var. botrytis cauliflower</td>
<td>FCM</td>
<td>Arumuganathan and Earle 1991</td>
</tr>
<tr>
<td></td>
<td>696</td>
<td></td>
<td></td>
<td>Johnston et al. 2005</td>
</tr>
<tr>
<td></td>
<td>752</td>
<td></td>
<td>FD</td>
<td>Bennet and Smith 1991</td>
</tr>
<tr>
<td></td>
<td>868</td>
<td></td>
<td>FD</td>
<td>Bennet and Smith 1976</td>
</tr>
<tr>
<td>B. juncea (AB)</td>
<td>1068</td>
<td>brown mustard</td>
<td>FCM</td>
<td>Johnston et al. 2005</td>
</tr>
<tr>
<td></td>
<td>1105</td>
<td></td>
<td>FCM</td>
<td>Arumuganathan and Earle 1991</td>
</tr>
<tr>
<td></td>
<td>1149</td>
<td></td>
<td>FD</td>
<td>Bennet and Smith 1976</td>
</tr>
<tr>
<td>B. napus (AC)</td>
<td>1129–1235</td>
<td>Rapeseed</td>
<td>FCM</td>
<td>Arumuganathan and Earle 1991</td>
</tr>
<tr>
<td></td>
<td>1127</td>
<td></td>
<td>FD</td>
<td>Bennet and Smith 1991</td>
</tr>
<tr>
<td></td>
<td>1132</td>
<td></td>
<td>FCM</td>
<td>Johnston et al. 2005</td>
</tr>
<tr>
<td>B. carinata (BC)</td>
<td>1284</td>
<td>Ethiopian mustard</td>
<td>FCM</td>
<td>Johnston et al. 2005</td>
</tr>
<tr>
<td></td>
<td>1544</td>
<td></td>
<td>FD</td>
<td>Bennet and Smith 1976</td>
</tr>
</tbody>
</table>

FCM, flow cytometry; FD, Feulgen densitometry.
karyotyping technique was further utilized by Olin-Fatih in a subsequent
work conducted on protoplasts (Olin-Fatih 1996). A study on root tips was
also conducted by Cheng et al. (1995a), who published a revised karyotype
of *B. oleracea* var. *alboglabra*. However, it should be noted that a relatively
weak confidence of chromosome identification precludes other researchers
from adopting the banding methods in their research as a standard protocol
for chromosome discrimination in *Brassica*. The karyotyping based on
G-banding applied by Olin-Fatih (1994) also cannot be accepted as a routine
way for this purpose. All the studies mentioned above show that in *Brassica*,
morphometric uniformity of chromosomes and their relatively high number
are the main limiting factors for efficient karyotype construction.

Like in the pachytene stage of meiosis, mitotic prometaphase
chromosomes exhibit an uneven pattern of chromatin condensation. This
fact has been utilized in measuring chromatin density along prometaphase
chromosomes (CP, condensation pattern = density profile) (Nakayama
and Fukui 1997). This method, together with 45S rDNA FISH-based
mapping, were further used for constructing karyotypes of three diploid
*Brassica* species (Fukui et al. 1998), and rapeseed (Kamisugi et al. 1998).
Nevertheless, “quantitative karyotyping” has not been widely adopted
in *Brassica* research, as it is relatively time-consuming and needs special
software equipment.

Later trials of karyotype construction utilized both chromosome length/
morphology, as well as 45S and 5S rDNA FISH analysis for chromosome
identification (Kulak et al. 2002; Snowdon et al. 2002). These attempts will
be further described in Section 6. In *B. oleracea*, the problem lies in a lower
number of chromosomes with 5S and 45S rDNA loci, which permits the
identification of only two out of nine chromosome pairs. More reliable
chromosome identification is achieved by using low-copy sequences,
especially derived from *Brassica* BAC libraries currently developed. A more
promising situation exists in *B. rapa*, where six out of 10 chromosome pairs
can be identified using 5S and 45S rDNA probes (Snowdon et al. 2002).
Koo et al. (2004) used 5S and 45S rDNA in combination with a 350-bp
centromeric tandem repeat, to distinguish both mitotic metaphase and
pachytene *B. rapa* chromosomes. Due to the specific hybridization pattern of
the centromeric repeat hybridizing to eight out of 10 chromosome pairs, the
construction of the *B. rapa* karyotype was greatly facilitated. The pachytene
karyotype described by Koo et al. (2004) comprises one acrocentric, two
metacentric, five submetacentric, and two subtelocentric chromosomes.
Moreover, the study gives us the opportunity to compare sizes of mitotic
metaphase *B. rapa* chromosomes with their pachytene counterparts
(Table 6-4). All *B. rapa* chromosomes are 15–20 times longer at the pachytene
stage as compared with mitotic metaphase chromosomes; similar to what
was described for *B. oleracea* (Ziolkowski and Sadowski 2002).
Armstrong et al. (1998) published an idiogram of *B. oleracea* var. *alboglabra* short-cycle line A12DHd, which was a parental genotype of one of the current *B. oleracea* mapping populations (Bohuon et al. 1996; Sebastian et al. 2000), and has been used in genome sequencing projects. Idiograms of mitotic prometa phase chromosomes were constructed and further used as a basis for more detailed studies, in which chromosome-specific probes were developed (Howell et al. 2002). Howell et al. (2002) used a number of genomic clones as FISH landmarks for individual chromosomes. These clones included a genomic restriction fragment length polymorphism (RFLP) clone, cDNA plasmid clone, cosmid clone, as well as 17 *B. oleracea* BACs. Selection of the BAC clones was based on BAC-filter hybridization with a number of known genomic sequences, and their assignment to specific loci was further confirmed by BAC clone fingerprinting, Southern hybridization, and locus-specific PCR (Howell et al. 2002). Most BACs contained, beside unique sequences, also DNA repeats (mainly TEs), which showed unspecific hybridization signals. Therefore, the inclusion of C0\textsubscript{r} blocking DNA was required to suppress the hybridization of repetitive sequences. Once having chromosome-specific landmarks at hand, emphasis was put on assignment of linkage groups to chromosomes. This assignment was done by applying a number of mapping strategies, including RFLP, cleaved amplified polymorphic sequences (CAPS) and single nucleotide polymorphism (SNP). Since there is more than one probe specific for each of the nine chromosomes of *B. oleracea*, these chromosomes can be oriented with respect to the linkage groups. The integrated cytogenetic and genetic maps of *B. oleracea* may be useful for estimation of the physical distance between markers applied in positional cloning. It should also be possible to estimate the physical length of a DNA region from a donor parent in a

<table>
<thead>
<tr>
<th>Chromosome no. (^a)</th>
<th>Chromosome length (µm ± SD) (^b)</th>
<th>Centromeric index (%±SD) (^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mitotic metaphase</td>
<td>pachytene</td>
</tr>
<tr>
<td>1</td>
<td>3.30±0.21</td>
<td>51.3±7.45</td>
</tr>
<tr>
<td>2</td>
<td>2.89±0.25</td>
<td>48.0±3.05</td>
</tr>
<tr>
<td>3</td>
<td>2.53±0.17</td>
<td>47.1±6.20</td>
</tr>
<tr>
<td>4</td>
<td>2.35±0.16</td>
<td>41.4±3.64</td>
</tr>
<tr>
<td>5</td>
<td>2.20±0.21</td>
<td>40.8±6.16</td>
</tr>
<tr>
<td>6</td>
<td>2.00±0.20</td>
<td>36.8±2.49</td>
</tr>
<tr>
<td>7</td>
<td>1.84±0.12</td>
<td>36.8±4.41</td>
</tr>
<tr>
<td>8</td>
<td>1.73±0.16</td>
<td>32.7±3.18</td>
</tr>
<tr>
<td>9</td>
<td>1.62±0.15</td>
<td>27.0±2.62</td>
</tr>
<tr>
<td>10</td>
<td>1.46±0.14</td>
<td>23.7±2.97</td>
</tr>
</tbody>
</table>

\(^a\)Chromosomes arranged according the decreasing length.

\(^b\)Centromeric index is the percentage of short arm by total chromosome length (Levan et al. 1964).

Table 6-4 A comparison of *B. rapa* subsp. *pekinessis* mitotic metaphase chromosomes with their pachytene counterparts (according to Koo et al. 2004).
substitution line using FISH probes associated with markers at each end of the substitution chromosome. But the most significant application of the integrated map lies in assignment of BACs of an overlapping set of clones (i.e. a BAC contig) to linkage groups for the construction of physical *Brassica* maps. Integrating genetic, cytogenetic and physical information coming from different research groups will lead to a much more comprehensive understanding of the *Brassica* genome architecture and its structural evolution.

6.6 Chromosomal Localization of rDNA in *Brassica*

The two types of ribosomal RNA genes, 5S rDNA and 45S rDNA encoding for 5S and 18S-5.8S-25S rRNAs, respectively, are widely used in FISH studies, providing convenient chromosome markers. As these two gene clusters are organized in head-to-tail tandem arrays localized on a few chromosomes in all known eukaryotic species, with each locus built of several hundred to a few thousand gene copies (Copenhaver and Pikaard 1996), they are easy to detect by the FISH technique.

Ribosomal RNAs, together with some protein complexes, build up ribosomes, the translation factory for protein synthesis. There are four classes of rRNA molecules in plant ribosomes, named 5S, 5.8S, 18S and 25S according to their Svedberg coefficient (see Pederson and Politz 2000, for review). Of these, the 5.8, 18 and 25S rRNAs result from splicing of a single 45S transcript encoded by rDNA repetitive units clustered at particular chromosomal loci. The fourth rRNA class, 5S rRNA, is transcribed from different gene arrays located usually in distinct chromosomal loci (Heslop-Harrison 2000). Transcription of the 45S rDNA gives rise to the most conspicuous morphological element within the nucleus, the nucleolus; hence, the loci where 5.8-18-25S rRNA genes are localized are termed nucleolar organizer regions (NORs) (McC1ntock 1934; Fig. 6-4). rRNA gene transcription is suspended during meiosis, which results in the disappearance of the nucleolus in the late meiotic prophase I in higher eukaryotes. The activity of the 45S rDNA loci at mitotic interphase can often be visualized at the following metaphase as a secondary constriction (McC1ntock 1934), thought to reflect the necessary DNA decondensation that allows the access of transcription machinery at interphase. At the loci of 45S rDNA, which were involved in nucleolus formation during interphase, some of the rRNA gene transcription proteins are retained at metaphase (Roussel et al. 1996). These proteins possess the ability to reduce silver under acid conditions, and this enables a differential staining of mitotic NORs (so called Ag-NORs) that were transcribed during the preceding interphase, as well as the staining of nucleoli in interphase nuclei (Goodpasture and Bloom 1975; Robert-Fortel et al. 1993). For the first time, this procedure was applied
to *Brassica* chromosomes by Cheng and Heneen (1995), which resulted in a much more precise description of NORs in dividing cells.

![Figure 6-4](image)

**Figure 6-4** Organization of 45S rDNA locus in plants. Genes are organized as head-to-tail repeats, each block consists of 18S-5.8S-25S rRNA gene units. Similar but simpler organization exists for the 5S rRNA gene cluster.

In *Brassica* cytogenetics, NORs were used as convenient markers distinguishing chromosomes (e.g., Röbbelen 1960; Olin-Fatih and Heneen 1992; Cheng et al. 1995a; Olin-Fatih 1996). Adaptation of the silver staining technique also contributed to the use of rDNA loci as convenient chromosomal markers (Cheng and Heneen 1995). On the other hand, by these means not all 45S rDNA loci can be identified, because only transcriptionally active genes contribute to the nucleolus formation. The progress in FISH techniques provided a much needed tool for the chromosome localization of rRNA genes.

Maluszynska and Heslop-Harrison (1993) were the first to successfully apply 45S rDNA as a molecular marker in FISH to *Brassica* chromosomes. The authors analyzed the presence of the 45S rDNA in diploid species (*B. nigra*, *B. oleracea*, and *B. rapa*), as well as in their allotetraploid derivates (*B. carinata*, *B. juncea*, and *B. napus*). They found two major pairs of rDNA loci in *B. nigra*, two major pairs and one minor pair in *B. oleracea*, and five pairs of loci in *B. rapa*. Among the three tetraploid species, *B. carinata* had four loci, *B. juncea* had five major pairs and one minor pair, and *B. napus* had six pairs of loci. These results indicated that the number of 45S rDNA loci has been reduced during the evolution of allotetraploid species. Later, Snowdon et al. (1997a) confirmed these results for *B. juncea* and *B. napus*. However, the authors found the number of rDNA sites in *B. carinata* equaled the sum of loci observed in the parental species. Furthermore, based on
the morphology of prometaphase chromosomes and rDNA hybridization patterns, Snowdon et al. (1997a) were able to assign individual B. napus rDNA loci to the parental A and C genomes, respectively. They proposed the reduction of rDNA loci number in the amphidiploid species resulting from the loss of the smallest site of the A genome (Snowdon et al. 1997a). The study of Fukui et al. (1998), based on the analysis of prometaphase chromosomes, was only partially consistent with the results obtained by Maluszynska and Heslop-Harrison (1993) and Snowdon et al. (1997a). The FISH experiments revealed five, three and two 45S rDNA loci in B. rapa, B. nigra and B. oleracea, respectively. The authors suggested that these differences may reflect an intraspecific polymorphism of rDNA loci.

In B. oleracea, the integration of genetic and physical chromosome maps started with the study of Armstrong et al. (1998), as described in Section 5. The authors used 5S and 45S rDNA as FISH probes, confirming the presence of three loci of 45S rDNA loci, previously reported for the species (Maluszynska and Heslop-Harrison 1993; Snowdon et al. 1997a). Using both mitotic and pachytene chromosomes, they found two 5S rDNA sites located on the long arm of chromosome O4 (chromosome numbering according to Howell et al. 2002). Later analysis, also performed on pachytene chromosomes, revealed that in fact there are three 5S rDNA loci on this chromosome arm (Ziolkowski and Sadowski 2002; Fig. 6-5).

Other studies used 5S and 45S rDNA probes for karyotype construction of the species of the U-triangle (Hasterok et al. 2001; Kulak et al. 2002; Snowdon et al. 2002). Although multicolour FISH (5S and 45S rDNA probes applied simultaneously) in combination with more extended prometaphase chromosomes results in more confident chromosome identification than with metaphase preparations, chromosome size and morphology still remain the main limiting factors in such approaches. Although a polymorphism in rDNA loci has been reported (see below), eight types of rDNA-bearing chromosomes in Brassica species of the U-triangle could be distinguished according to Hasterok et al. (2001) and Kulak et al. (2002) (Fig. 6-5).

6.6.1 Nucleolar Dominance in Brassica

An interesting epigenetic phenomenon observed for genes encoding for 18S-5.8S-25S rRNA is nucleolar dominance. It is based on transcriptional dominance of the ribosomal genes from one species over the genes from the second species in interspecific hybrids.

Both molecular and cytogenetic approaches were used to analyze the nucleolar dominance in Brassica species. For cytogenetic studies, the silver staining method (AgNOR; for principles of the technique, see above) could be applied, as this technique enables detection of only transcriptionally active 45S rRNA genes (Goodpasture and Bloom 1975). These studies were
conducted in Brassicas from the 1990s, when the nucleolar activity of 45S rDNA loci was described for three diploid *Brassica* species (Cheng and Heneen 1995; Cheng et al. 1995b). Combining FISH with AgNOR facilitated discrimination of active and inactive loci in *Brassica* amphidiploids (Hasterok and Maluszynska 2000b). This study showed that although the number of loci in amphidiploids is reduced, the number of active loci equals to the number of AgNORs in their diploid ancestors. Active rDNA loci are located at the secondary constrictions, while loci at pericentromeric or terminal locations are inactive or their activity is below detection. More interestingly, these findings indicate that there is no nucleolar dominance in root meristematic cells of *Brassica* species. However, the result seems to be contradictory to the molecular studies of Chen and Pikaard (1997b), who presented convincing evidence for nucleolar dominance in young leaves of allotetraploid *Brassica* species. The discrepancy can be explained by different patterns of rRNA gene expression in various organs and developmental stages, which was partially confirmed by other molecular studies (Chen and Pikaard 1997a).

Molecular studies elucidated the epigenetic nature of nucleolar dominance in amphidiploid *Brassica* species (Chen and Pikaard 1997a). It was shown that nucleolar dominance can be reversed by chemical inhibitors.
of DNA methylation or histone deacetylation. These findings led to the
conclusion that nucleolar dominance is a consequence of the selective
silencing of rRNA genes from one progenitor rather than the selective
activation of the other (Chen and Pikaard 1997a). Further studies carried
out on Arabidopsis plants have shown that small interfering RNA (siRNA)-
directed DNA methylation and changes in posttranslational histone
modifications (mainly acetylation and methylation) are responsible for the
“on-off switch mechanism” controlling the number of active rRNA genes
(Lawrence et al. 2004; Preuss et al. 2008). Moreover, the mechanism seems
to be the same both for interspecific hybrids exhibiting nucleolar dominance
and for non-hybrid, diploid plants, where it controls the effective dosage
of rRNA in response to the particular physiological demands of the cell
(Lawrence et al. 2004).

6.6.2 Evolutionary Trends in rDNA Loci Number Change

One of the most intriguing aspects of the evolution of rDNA in Brassica
species is connected with the differences in the number of rDNA loci found
among closely related species, their intraspecific variation, and increase/
decrease of their number in Brassica amphidiploids.

First clues of rDNA copy number variation came from molecular
analyses. Restriction fragment length polymorphism (RFLP) applied to
B. rapa-B.oleracea addition lines revealed two loci of 45S rDNA (Delseny
et al. 1990). A similar approach applied on dissected B. oleracea chromosomes
(using monosomic addition lines) revealed strong rDNA hybridization
signals linked to two linkage groups, and a weak signal linked to a third
linkage group (McGrath et al. 1990). Further studies on the segregation
of rDNA RFLP markers in cross progenies have indicated the presence
of three pairs of rRNA gene loci in one of the four B. oleracea populations
investigated (Kianian and Quiros 1992a, b).

These initial investigations conducted in B. oleracea only indicated the
phenomenon of a rapid rDNA evolution in the Brassica genus. Much more
data were collected based on FISH analyses. One of the first unexpected
results was a finding that the number of 45S rDNA loci in allotetraploids
is reduced in comparison with their diploid ancestors (Maluszynska and
Heslop-Harrison 1993; Snowdon et al. 1997a). Reduction of 45S rDNA loci
number was also observed in some other polyploid species and genera,
including Nicotiana tabacum (Volkov et al. 1999), and rice (Shishido et al.
2000).

It is still not clear what genetic mechanisms trigger rapid evolution
of 45S rDNA loci. However, comparative mapping in the Brassicaceae
family shows that most of the chromosome breakages have occurred close
to centromeres or telomeres, where long stretches of tandem repeats exist
(Boivin et al. 2004; Kuittinen et al. 2004; Yogeeswaran et al. 2005; Henry et al. 2006; Lysak et al. 2006). As 45S rDNA loci are found usually at chromosome ends or within pericentromeric regions, it may imply that their extremely high mobility within the genome and frequent loss in hybrids is a result of their breakage-prone location (Schubert 2007). In addition, 45S rDNA tandem repeats are prone to recombination and rearrangements.

The reduction of loci number after polyploid formation is also observed in the case of 5S rRNA genes. A loss of rDNA loci has apparently occurred in some polyploid species within the Brassicaceae family, as in polyploid species *Moricandia arvensis* (2n = 28), and *Alyssum spinosum* (2n = 32), only one 5S rDNA locus was reported (Ali et al. 2005). Reduction of loci number has been also found in *B. carinata* (both parental species have six 5S rDNA loci, while the polyploid only four; see Table 6-5). Because of the documented large numbers of 5S rDNA loci and often contradictory reports, the question of 5S rDNA loci decrease/increase in *B. napus* and *B. juncea* could not be resolved as yet (see Table 6.5 for references).

The phenomenon of reduction of rDNA loci in polyploids is usually explained as one of the first steps in genome diploidization, the process by which a polyploid organism attains the diploid state (Wendel 2000). After hybridization of genomes derived from two different species, an imbalance in gene dosage can prevail, which leads to inadequate amounts of different gene products in cells. In the case of rDNA genes, an excess of ribosomal RNA produced by multiplied rDNA loci can further cause disturbances.

### Table 6-5 Chromosome localization of 5S and 45S rRNA genes in *Brassica* species.

<table>
<thead>
<tr>
<th>Species</th>
<th>Genome(s)</th>
<th>No. of chromosomes with 5S rDNA loci</th>
<th>No. of chromosomes with 45S rDNA loci</th>
<th>No. of chromosomes with 5S and 45S rDNA loci</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. nigra</em></td>
<td>BB</td>
<td>2 (t)</td>
<td>6</td>
<td>1,4,6,9,13,14</td>
<td></td>
</tr>
<tr>
<td><em>B. oleracea</em></td>
<td>CC</td>
<td>4* (4,6)*** (i)</td>
<td>4+2 minor (i)</td>
<td>1,2,3,4,6,9,10,12,14</td>
<td></td>
</tr>
<tr>
<td><em>B. rapa</em></td>
<td>AA</td>
<td>6 (6-10)**</td>
<td>10 (i) (6,10)***</td>
<td>1,2,4,6,9,12,14</td>
<td></td>
</tr>
<tr>
<td><em>B. carinata</em></td>
<td>BBCC</td>
<td>4 (i)</td>
<td>8 (8-10)**</td>
<td>1,6,9,11,14</td>
<td></td>
</tr>
<tr>
<td><em>B. juncea</em></td>
<td>AABB</td>
<td>10 (8,10)**</td>
<td>12-16 (i)</td>
<td>1,6,9,11,14</td>
<td></td>
</tr>
<tr>
<td><em>B. napus</em></td>
<td>ACCC</td>
<td>14* (10,12)**</td>
<td>12-14 (i)</td>
<td>1,2,5,6,7,8,9,10,11,12,14</td>
<td></td>
</tr>
</tbody>
</table>


*Two 5S rDNA sites on the same chromosome.

** Polymorphism of loci number reported by Hasterok et al. (2006).

*** Polymorphism of loci number reported by Ziolkowski and Sadowski (2002).

(i), interstitial positions of rDNA exclusively; (t), terminal/subterminal positions of rDNA exclusively.
in the ribosome function. For this reason, reduction of unnecessary 5S or 45S rDNA can be beneficial for the newly formed polyploid (Dvorak 1990; Wendel 2000). However, opposite trends during polyploid formation are sometimes observed, as shown in cotton (Hanson et al. 1996) and *Sanguisorba* species (Mishima et al. 2002). In those species the trend towards the increase of rDNA loci number in newly formed polyploids can be explained in an analogous way, by a requirement to compensate a deficiency in rRNA production. This could be true especially in the cases, in which a polyploid plant tends to be larger and demonstrates a more rapid growth than its progenitor(s). However, spontaneous chromosomal rearrangements inside and between the rRNA gene loci cannot be also ruled out.

Comparative studies within the *Brassica* genus revealed that polymorphism in the number and chromosomal position of rDNA occurs among various subspecies, varieties, and even within a population of one cultivar or variety (Hasterok et al. 2006). A high level of polymorphism was especially found in the species containing the A genome, i.e. *B. rapa*, *B. juncea*, and *B. napus*, and it is supposed that this phenomenon is linked to extraordinary high numbers of both 5S and 45S rDNA loci in the A genome (Hasterok et al. 2006). The polymorphism observed may reflect differences in transcriptional activity of a particular rRNA gene loci among different forms of the species. These differences can be indirectly inferred from variation in length of secondary constrictions. Other examples of polymorphism are due to the significant variation in copy number of tandemly arrayed rRNA genes in a locus, e.g., one to three copies reported for 5S rDNA loci in *B. oleracea* (Hasterok et al. 2006 for a review).

Interesting conclusions about 5S rDNA loci spreading in *Brassica* can be drawn from analysis of triplicated chromosomal block N (Ziolkowski et al. 2006; see also below). This chromosomal region is unique in the *A. thaliana* genome, but occurs in three copies in diploid *Brassica* genomes as a consequence of a whole-genome triplication (Lysak et al. 2005; Parkin et al. 2005; Ziolkowski et al. 2006). One of the three genomic copies in the *B. oleracea* genome contains two loci of 5S rRNA genes, at interstitial location (Ziolkowski et al. 2006). The whole block is close to another pericentromeric 5S rDNA locus, which is considered to be ancestral, based on the interspecies comparison (Hasterok et al. 2006). A most probable explanation of such a pattern is a small-scale transposition of the 5S rDNA locus from its original pericentromeric location to two sites within the adjacent block N by an unknown mechanism, possibly by rDNA-containing extrachromosomal circular DNA (eccDNA; Cohen et al. 2008). Further expansion of copy number was probably the result of unequal crossing-over (Fig. 6-6).
6.7 Genome-Specific Repeats and Genomic in Situ Hybridization

That both tandem and dispersed TE repetitive sequences can exhibit a very variable level of the sequence similarity among inter-related taxa. DNA repeats shared among a group of species usually underline their close relatedness and common ancestry. On the contrary, the dynamic evolution and/or long-term genetic isolation of repetitive elements results in their sequence divergence, and the origin of genome- and taxon-specific repeats. Both opposed phenomena mark the evolution of repetitive elements in *Brassica* and are being used in cytogenetic and phylogenomic analyses of this genus (Lysak and Lexer 2006).

Genome-specific dispersed repetitive elements allow for a specific discrimination of individual genomes or chromosomes within composite genomes of allopolyploid species, interspecific hybrids, and introgression lines. In GISH experiments, typically genomic DNA (gDNA) of one genome/species is fluorescently labeled and hybridized to chromosomes of the composite genome in an excess of unlabeled gDNA or repeat-enriched \((C_{r-1})\) DNA fraction of the second parental genome. The unlabeled gDNA hybridizes with common repeats in the probe gDNA, minimizing the potential intergenomic cross-hybridization. Alternatively, both parental
gDNAs can be differentially labeled and hybridized simultaneously without the addition of blocking DNA.

As three important Brassica species (B. carinata, B. juncea, B. napus) have an allopolyploid origin (U 1935), GISH has been used frequently to discern parental genomes or individual chromosomes in these species, their derivates, and intergeneric hybrids between B. napus and other crucifer species. Parental genomes (A, B, and C) can be identified using GISH in the allotetraploid genomes of B. carinata (BBCC) and B. juncea (AABB) (Snowdon et al. 1997b; Hasterok et al. 2005a; Maluszynska and Hasterok 2005). However, donor genomes of B. napus (AACC) could not be clearly distinguished (Snowdon et al. 1997b; Hasterok et al. 2005a) due to the close phylogenetic relationship between B. oleracea (CC) and B. rapa (AA) (e.g., Lysak et al. 2005; Warwick and Sauder 2005). Contrary to these results, recently the C genome has been unequivocally identified in B. napus using labeled B. oleracea as a probe and approximately a 10-fold excess of unlabeled blocking DNA of B. rapa (Howell et al. 2008; Fig. 6-7). The reverse combination did not yield convincing results as the signals of labeled B. rapa DNA were overall weaker and cross-hybridization to pericentromic regions of B. oleracea-derived chromosomes has been observed. As crucifer species including Brassica species possess generally small genomes and correspondingly low amount of genome-specific repetitive elements (Lysak et al. 2009), GISH has a limited application in Brassica cytogenetics.

![Figure 6-7 GISH in Brassica napus. GISH with labeled genomic DNA of B. oleracea (red) revealed 18 mitotic chromosomes corresponding to the C genome. Unlabeled DNA of B. rapa (A genome) and the intergenic spacer region of B. oleracea 45S rDNA were used as a blocking DNA. Chromosomes counterstained with DAPI. Courtesy of S. J. Armstrong (University of Birmingham).](image_url)

*Color image of this figure appears in the color plate section at the end of the book.*
specific repeats are primarily clustered within pericentromeric regions, and therefore chromosome arms remain frequently unlabeled and potential intergenomic translocations undetected (e.g., Snowdon et al. 1997b; Wang et al. 2004). This spatially limited labeling was also observed by Howell et al. (2008) on A genome chromosomes in *B. napus* when gDNA of *B. rapa* has been used as a probe. Conversely, gDNA of *B. oleracea* labeled corresponding C-genome chromosomes homogeneously (see Fig. 6-7) to such an extent that a reciprocal translocation between A genome and C genome chromosomes could be detected (Howell et al. 2008). Different labeling efficiency of the two gDNA probes also suggests that *B. oleracea* chromosomes contain more and equally distributed genome-specific repetitive sequences than found in the chromosomes of *B. rapa*.

GISH was also extensively used to identify parental chromosomes in intergeneric hybrids and addition lines of *B. napus* and a number of other crucifer species including *A. thaliana* (Leino et al. 2004), *Crambe abyssinica* (Wang et al. 2004), *Eruca sativa* (Fahleson et al. 1997), *Lesquerella fendleri* (Skarzhinskaya et al. 1998), *Orychophragmus violaceus* (Hua et al. 2006; Zhao et al. 2008 and references therein), *Raphanus sativus* (Snowdon et al. 1997b), *R. raphanistrum* (Benabdelmouna et al. 2003), *Sinapis arvensis* (Snowdon et al. 2000a), and *S. alba* (Wang et al. 2005).

6.8 FISH with Repetitive Sequences Other Than rDNA

The chromosome organization of repetitive elements does not have only an important evolutionary significance but diverse classes of *Brassica* repeats serve genome and chromosome identification. In *Brassica* genomes, heterochromatic pericentromeric and centromeric regions are dominated by tandem (satellite) repeats, whereas transposable elements reside in these regions as well as along largely euchromatic chromosome arms.

Similar to other crucifer species (Lysak and Lexer 2006), satellite repeats are an abundant component of *Brassica* centromeres. In *B. rapa*, two 176-bp centromeric satellite repeats sharing ca. 80% sequence homology have been identified (Harrison and Heslop-Harrison 1995; Koo et al. 2004; Lim et al. 2005). The CentBr1 and CentBr2 satellites reside at eight and two centromeres, respectively (Lim et al. 2005). Both CentBr1 and CentBr2 were also found in *B. oleracea*, with the former and latter repeat localized to all and at least five centromeres, respectively, but not detected in the *B. nigra* genome (Lim et al. 2007). The expected number of chromosomes with CentBr1 and CentBr2 repeats corresponding to the donor (A and C) genomes was identified in *B. carinata*, *B. juncea*, and *B. napus* (Lim et al. 2007). In *B. nigra*, a 329-bp tandem repeat was shown to be specific for the B genome, and confirmed also on 16 B-genome chromosomes in the allotetraploid *B. juncea* (Schelfhout et al. 2004). The genome specificity of centromeric satellites further underlines
closer phylogenetic relationship between A and C genomes (B. rapa and B. oleracea) as compared to the B genome (B. nigra) (Lysak et al. 2005; Warwick and Sauder 2005). Distal parts of Brassica chromosomes were explored much less than (peri)centromeric regions, and except the Arabidopsis-type telomeric sequence (Hasterok et al. 2005a), only a single telomere-like repeat has been in situ localized on Brassica chromosomes (dos Santos et al. 2007). The clone pBo1.6 isolated from B. oleracea was localized at chromosome termini as well as presumably interstitial positions in B. oleracea, whereas almost no hybridization to B. rapa chromosomes has been observed. In B. napus, the repeat labeled all C genome chromosomes and interestingly up to six chromosomes of the A genome.

Several TEs were identified in Brassica species and localized on chromosomes. Alix and Heslop-Harrison (2004) isolated a number of retrotransposons from six species of the U’s triangle, and analyzed chromosomal distribution of two Ty1-copia-like and one Ty3-gypsy-like retrotransposons in B. oleracea and B. rapa (Alix et al. 2005). One Ty1-copia probe hybridized along the full chromosome length with some interstitial clustering, whereas another copia element, one Ty3-gypsy probe, and an Athila-like gypsy element hybridized along entire chromosomes with stronger signals at (peri)centromeric regions. Centromere- and pericentromere-specific LTR retrotransposons of B. rapa were analyzed by Lim et al. (2007). The CRBs (centromeric retrotransposons of Brassica) were localized as nested within the centromeric CentBr satellite repeat in all species of the U’s triangle. Peri-centromere retrotransposons of B. rapa (PCRBr) are Ty3-gypsy-like elements specific for the A genome (localized on three B. rapa chromosomes) and absent in the B and C genomes. Surprisingly, the PCRBr probe hybridized to B and C genome chromosomes in Brassica allotetraploids (Lim et al. 2007). Alix et al. (2008) isolated and characterized a CACTA transposon Bot1 (B. oleracea transposon 1) comprising ca. 2.3% of the B. oleracea genome. The Bot1 is a C genome-specific transposon amplified in the genome of B. oleracea but not in B. rapa. In B. napus, the Bot1 used as a FISH probe specifically labels 18 chromosomes of the C genome with only weak hybridization to B. rapa-derived chromosomes (Fig. 6-8). The C genome specificity and homogeneous chromosome distribution of this transposon can be probably linked with uniform GISH labeling pattern of the B. oleracea gDNA probe in B. napus (Howell et al. 2008; Fig. 6-6). Bot1 has been found to be useful for cytogenetic identification of A and C genome chromosomes, respectively. A B. oleracea BAC clone containing Bot1 was used to identify C genome chromosomes in inter-species AAC triploid hybrids (Leflon et al. 2006) and in B. napus haploids (Nicolas et al. 2007). Similarly, a B. rapa BAC containing an unknown repetitive sequences labeled specifically three pairs of A genome chromosomes in AAC hybrids (Leflon et al. 2006).
The specific structure of crucifer genomes with most repeats clustered in discrete heterochromatic arrays and well-developed genomic resources including chromosome-specific libraries of high-capacity DNA vectors were the main essentials for establishing BAC FISH and chromosome painting in Brassicaceae. There are several types of large-insert DNA vectors available [BACs, yeast artificial chromosomes (YACs), P1-derived artificial chromosomes (PACs), and transformation-competent artificial chromosomes (TACs)], however, only bacterial artificial chromosomes (BACs) became widely used as DNA probes in plant cytogenetics. Without making a clear-cut distinction, BAC FISH generally refers to in situ localization of individual or several BAC clones, whereas chromosome painting applies to visualization of large chromosome regions such as chromosome arms or whole chromosomes using BAC contigs, i.e., continuous sequences of overlapping or non-overlapping BACs.

In *Brassica* cytogenetics, BAC FISH has been mainly used for chromosome identification and integration of genetic and cytogenetic maps. Several BAC libraries of three *Brassica* species (*B. oleracea*, *B. rapa*, and *B. napus*) are available (see The Multinational *Brassica* Genome Project, http://www.brassica.info). BAC clones of the two diploid species have been applied most frequently as cytogenetic markers. Howell et al. (2002) assigned nine genetic linkage groups of *B. oleracea* to corresponding mitotic chromosomes using chromosome-specific BAC clones (see Section 5 for details). More
recently, FISH of three chromosome-specific BACs followed by GISH was used to characterize a reciprocal translocation in *B. oleracea* (Howell et al. 2008). BAC FISH has been extensively applied in the integration of genetic and cytogenetic maps within the *Brassica rapa* Genome Project (http://www.brassica-rapa.org). Chromosome/linkage-specific *B. rapa* BACs have been identified by unique expressed sequence tag (EST) clones and other genetic markers, and mapped to pachytene chromosomes by FISH (Yang et al. 2005; Lim et al. 2006). *B. rapa* BACs can also be localized on mitotic metaphase chromosomes as shown for BACs containing paralogs of the *FLOWERING LOCUS C* gene (Yang et al. 2006). In on-going *Brassica* sequencing projects, BAC FISH will be used as an important tool anchoring the sequenced clones and contigs on chromosomes (see also Chapter 7).

Comparative cytogenetic studies of non-repetitive chromosome regions are based on cross-species FISH of BACs derived from one species to chromosomes of another related species. In cytogenetics of Brassicas, *Arabidopsis* BACs and BAC contigs were exclusively used to reveal cross-species chromosome homeology. Comparative chromosome painting (CCP) and BAC FISH in *Brassica* species serve practical as well as more theoretical objectives. FISH with *Arabidopsis* BAC clones revealing homeologous chromosome regions in *Brassica* genomes can expedite the chromosome landing of *Brassica* BACs and the construction of cytogenetic maps. As large-scale CCP discerns the extent of interspecies chromosome homeology, valuable insights into karyotype and whole-genome evolution can be gained from such analyses. The first comparative BAC FISH in *Brassica* has been shown by Jackson et al. (2000) when a short contig of six *Arabidopsis* BACs was localized on DNA fibers and four to six mitotic chromosomes of *B. rapa*. Hybridization of *Arabidopsis* BAC clones to multiple chromosomes was also observed in *B. oleracea* as two partly overlapping *Arabidopsis* BACs hybridized to two different bivalents (Ziolkowski and Sadowski 2002), and 12 BAC clones were found segmentally duplicated on chromosome O6 (Howell et al. 2005). All three studies suggested that genomes of both the *Brassica* species were duplicated as compared to the genome of *A. thaliana*. Nevertheless, only three other large-scale CCP studies demonstrated the nature of the whole-genome duplication characterizing *Brassica* species. Lysak et al. (2005) analyzed the structure of a ~8-Mb *Arabidopsis* BAC contig in 21 species classified as members of the tribe Brassiceae including the six species of the U-triangle. As most species analyzed possessed three or six homeologous copies of the *Arabidopsis* contig, the data indicated that the *Brassica* genome duplication has to be specified as a whole-genome triplication proposed earlier by Lagercrantz (1998) and more recently by Parkin et al. (2005). In agreement with the phylogenetic analysis, revealed
chromosome homeologies advocated the existence of a common hexaploid ancestor for the whole tribe Brassiceae. The 8 to 15 million years old whole-genome triplication has been followed by genome diploidization manifested by chromosome number reduction accompanied by various chromosome rearrangements (Lysak et al. 2005). These findings have been corroborated by a CCP analysis of another *Arabidopsis* BAC contig in 10 species traditionally treated as members of Brassiceae (Lysak et al. 2007). Three copies of the analyzed chromosome segment found in the Brassiceae species further supported the hexaploid ancestry and monophyletic origin of the tribe. Ziolkowski et al. (2006) analyzed conservation of ancient rearrangement breakpoints between *A. thaliana* and *B. oleracea* genomes using three *Arabidopsis* BAC contigs spanning a total of 8.25 Mb. Also this work confirmed the triplication event preceding the species divergence in *Brassica* and Brassiceae (Fig. 6-9). Further CCP studies are discussed in Section 6.11.

![Figure 6-9](image)

**Figure 6-9** An example of comparative chromosome painting in *B. oleracea*. A 5.4 Mb long *Arabidopsis* BAC contig from chromosome 3 divided into six differentially labeled segments, was hybridized to pachytene chromosomes of *A. thaliana* (right) and *B. oleracea* (left). Whilst the BAC contig reveals a single chromosome region in *A. thaliana*, three copies of the contig were found in *B. oleracea* (arrowheads). Bioinformatic and genetic mapping studies suggest that the arrangement of A-F segments observed in *A. thaliana* has ancestral-like structure [(Henry et al. 2006, also data from http://chibba.agtec.uga.edu/duplication/index/home (Tang et al. 2008)]. This is observed in one of the three *B. oleracea* copies, as well. However, segment A was separated from segments B-F in the two other copies (dissociated A segments indicated by arrows). Modified from Ziolkowski et al. (2006).

*Color image of this figure appears in the color plate section at the end of the book.*
6.10 Whole-Genome Duplications in the Evolution of Crucifer Genomes

In the last decade the major progress in studies on genome evolution of Brassicaceae species was enabled due to the *Arabidopsis* genome sequencing project (AGI 2000). One of the most important reasons for choosing *A. thaliana* as a model plant species was a belief that its small genome (~157 Mbp; Bennett et al. 2003) built from five chromosomes, has a typical diploid-like structure. Thus, an intriguing result of genome sequencing was the discovery that the *Arabidopsis* genome is a mosaic of duplicated chromosomal segments (AGI 2000; Blanc et al. 2000; Paterson et al. 2000; Vision et al. 2000). This initial finding has led to an upsurge in bioinformatic studies uncovering the ancient evolutionary history of the *Arabidopsis* genome. At present, new data from genome sequencing projects in *Oryza sativa* (rice), *Populus trichocarpa* (poplar), *Vitis vinifera* (grapevine), *Carica papaya* (papaya) and *Sorghum bicolor* (sorghum) complement the data on the *Arabidopsis* genome. Such studies are based mainly on finding and comparing collinear regions (chromosomal regions with a similar gene content and order) within a genome and/or between different genomes. As it is unlikely that a pair of collinear regions could originate independently, it is assumed that it has a common origin, and was derived from independent segmental duplications and/or whole-genome duplication (WGD) events. In plants, WGDs are much more widespread than large segmental duplications, and played a pivotal role in the course of evolution (Adams and Wendel 2005). Most evolutionary events uncovered by mining the *Arabidopsis* genome are shared also by Brassica species, as they predate the divergence of the *Arabidopsis* and *Brassica* lineages.

Currently it is widely accepted that there were three WGDs during the evolution of the *Arabidopsis* genome (named γ, β, and α, or 1R, 2R and 3R, starting from the oldest one; Simillion et al. 2002, Blanc et al. 2003, Bowers et al. 2003). None of these events can be detected cytogenetically (i.e. by BAC FISH or chromosome painting) due to the significant structural divergence of duplicated chromosomal regions (Lysak et al. 2001). Recent comparisons of the *Arabidopsis* and poplar genome sequences with those of grapevine and papaya (Jaillon et al. 2007; Ming et al. 2008) demonstrated that the γ WGD was a hexaploidization event, and took place in a common ancestor of *A. thaliana*, *C. papaya*, *P. trichocarpa*, and *V. vinifera* (Jaillon et al. 2007). A similar comparison of sequence data from rice and sorghum showed that the γ event occurred in a lineage leading to Rosids after the separation from grasses (Jaillon et al. 2007; Paterson et al. 2009; Fig. 6-10).

Comparisons of the *Arabidopsis* genome architecture with those of poplar, papaya and grapevine also shed more light on the rate of chromosome rearrangements within the Brassicaceae lineage. It can be
concluded that the extant *Arabidopsis* genome, and hence all the *Brassica* genomes, are much more rearranged, than those of *V. vinifera* and *C. papaya*, while the *P. trichocarpa* genome represents an intermediate situation (Jaillon et al. 2007; Tang et al. 2008). At the level of gene collinearity, only remnants of the ancestral gene content and order could be identified in the *Arabidopsis* genome. Moreover, not only the frequency of rearrangements was significantly higher in the Brassicaceae ancestor (Tang et al. 2008), but also the gene content and order were much more eroded by a frequent gene loss, than in grapevine and papaya (Thomas et al. 2006). Both these phenomena are due to additional rounds of WGD followed by extensive diploidization process in the Brassicaceae family (Adams and Wendel 2005).

![Phylogenetic Tree](image)

*Figure 6-10* Described WGD events positioned on a phylogenetic tree in the evolution of selected dicotyledonous plants. WGDs are represented as stars and refer to the hexaploidization $\gamma$ shared by all Rosids, WGDs $\alpha$ and $\beta$ within the Brassicales, the salicoid duplication $p$ in the poplar lineage, and hexaploidization events $c$ and $b$ in Cleomaceae and Brassiceae, respectively. Branch lengths do not correspond to evolutionary distances.

Two more recent WGDs ($\beta$ and $\alpha$) described in *Arabidopsis* were presumably tetraploidizations (Blanc et al. 2003; Bowers et al. 2003; Simillion et al. 2003; Fig. 6-10). Recent intergenomic comparisons suggest that both the WGD events took place in Brassicaceae after the separation from the papaya lineage (Ming et al. 2008). Relatively well conserved gene collinearity between segments of $\beta$ and $\alpha$ duplications suggests that both events occurred within a short interval (Ziolkowski et al. 2003). It has been estimated that $\beta$ and $\alpha$ duplications cover 52% and 89% of the genome, respectively (Bowers et al. 2003), and only pericentromeric regions were not found to be duplicated, being too eroded by the activity of TEs (Blanc et al. 2003).
Several approaches have been applied for dating β and α WGDs (e.g., Blanc et al. 2003; Bowers et al. 2003; Maere et al. 2005; Schranz and Mitchell-Olds 2006). Most estimates were based on a comparison of the level of synonymous (silent) substitutions between genes duplicated through successive WGDs in Arabidopsis, and also with orthologous genes (homologous genes that originated through speciation) from other plant species. Both β and α events took place after the Arabidopsis–papaya divergence (~72 million years ago (Mya); Wikström et al. 2001), as there is no evidence for the two events in C. papaya (Ming et al. 2008). More research is needed to establish the position of the β WGD within Brassicales. On the other hand, data on the position of the α event is more accurate, as it was established that the event occurred shortly after the divergence of two sister families, Brassicaceae and Cleomaceae (represented by Cleome spinosa; Schranz and Mitchell-Olds 2006), some 24 to 40 Mya (Blanc et al. 2003, 2004). Therefore, it has been suggested that the α event is specific for Brassicaceae, however, more effort is needed to determine when the event occurred during the evolution of the family (Schranz and Mitchell-Olds 2006; Fig. 6-10). A detailed investigation of the genus Aethionema, which diverged early from the rest of the family (Hall et al. 2002) seems to be of major importance (Schranz and Mitchell-Olds 2006).

Even early cytogenetic studies suggested that present-day diploid Brassica genomes exhibit intragenomic chromosome homeology and duplications, and thus represent “balanced secondary polyploids” (e.g., Catcheside 1934; Thompson 1956; Röbbelen 1960). Comparative genetic mapping among B. nigra, B. oleracea and B. rapa (Lagercrantz and Lydiate 1996), and between A. thaliana and B. oleracea suggested an ancient hexaploidization as a cause of the intragenomic duplications observed in Brassica genomes (Lagercrantz 1998; Babula et al. 2003). An alternative hypothesis referred to as “cyclic amphiploidy” proposed that highly duplicated Brassica genomes originated from various hybridizations among different species with a genome based on \( x = 4 \) and/or \( x = 5 \) (Truco et al. 1996; Quiros 1999). However, a wealth of data coming from different research fields speaks for the hexaploidy hypothesis (i.e., Cavell et al. 1998; O’Neill and Bancroft 2000; Babula et al. 2003; Rana et al. 2004; Lysak et al. 2005, 2007; Parkin et al. 2005; Yang et al. 2005; Town et al. 2006; Ziolkowski et al. 2006). The whole-genome triplication has been also corroborated by comparative chromosome painting with large Arabidopsis BAC contigs as discussed in Section 6.9 (Lysak et al. 2005, 2007; Ziolkowski et al. 2006; Fig. 6-9). Moreover, CCP analysis of a number of Brassiceae species revealed that the polyploidization event is common for the whole tribe, and may serve as an apomorphous character defining the tribe (Lysak et al. 2005).
6.11 Chromosome Evolution in Brassiceae and Brassicaceae

To reconstruct the chromosomal evolution across the Brassicaceae, it is crucial to distinguish between derived and ancestral genome organization. Gene-based phylogenies enable us to estimate evolutionary distances and divergence times between individual species and species groups, and in consequence help to position genome alteration events to particular nodes within a phylogenetic tree. The 338 genera of the Brassicaceae family have been recently classified into 25 tribes based on nuclear- and chloroplast-encoded markers (Al-Shehbaz et al. 2006; Bailey et al. 2006; Beilstein et al. 2006, 2008). Sixteen of the 25 tribes can be further grouped into one of three major crucifer lineages (referred to as Lineages I-III; Al-Shehbaz et al. 2006; Beilstein et al. 2006, 2008). This treatment provides the most up to date reference for comparative studies in Brassicaceae, although it is being continuously modified as new data become available (e.g. German et al. 2008).

Chromosomal studies in Brassicaceae are hampered due to the high frequency of polyploidy. Polyploidy has played an important role in the evolution of the Brassicaceae family (Marhold and Lihova 2006), and nearly 37% of the species are polyploid (Warwick and Al-Shehbaz 2006). Up to date chromosome numbers have been reported for more than 1,500 crucifer species (summarized by Warwick and Al-Shehbaz 2006). Although a continuous series of base chromosome numbers from \( x = 4 \) to \( x = 13(17) \) exists, a surprisingly high percentage (37%) of genera appears to be based on \( x = 8 \) (Warwick and Al-Shehbaz 2006). This finding led to a hypothesis that an ancestral chromosome number in the family, and particularly in Lineage I, was \( n = 8 \) (Koch et al. 1999; Henry et al. 2006; Lysak et al. 2006; Schranz et al. 2006).

Investigation of the chromosomal evolution within the tribe Brassiceae has progressed significantly due to the comparisons with the \textit{Arabidopsis} genome \((n = 5)\). \textit{A. thaliana} is the best studied plant genome furnished with a wealth of genomic resources, and close enough phylogenetically to the Brassicaceae. However, it is now well documented that the \textit{Arabidopsis} genome experienced reshuffling coupled with chromosome number reduction from an ancestral \( n = 8 \) genome (Koch et al. 1999; Henry et al. 2006; Lysak et al. 2006). The reshuffled nature of the \textit{Arabidopsis} genome makes any interspecies comparison difficult as it must account for these species-specific rearrangements (Schranz et al. 2006). To facilitate further intrafamily studies, a tentative Ancestral Crucifer Karyotype (ACK) has been recently proposed (Lysak et al. 2006; Schranz et al. 2006).

The ACK concept has been based on comparative genetic maps of \textit{Arabidopsis lyrata} and \textit{Capsella rubella} (both \( n = 8 \)) constructed using \textit{Arabidopsis} genetic markers (Boivin et al. 2004; Kuittinen et al. 2004; Yogeeswaran et al. 2005). This three-way comparison showed that both \textit{A. lyrata} and
C. rubella possess genomes largely collinear with that of A. thaliana, and all three species share large conserved chromosomal blocks (Boivin et al. 2004; Kuittinen et al. 2004; Yogeeswaran et al. 2005). Moreover, genomes of A. lyrata and C. rubella are almost identical with respect to the gene order (Koch and Kiefer 2005), and thus are believed to resemble an ancestral \( n = 8 \) genome (Lysak et al. 2006; Schranz et al. 2006). This hypothesis was further tested by CCP approach (Lysak et al. 2006), when Arabidopsis BAC contigs were arranged according to the genetic map of A. lyrata/C. rubella, fluorescently labeled and hybridized to pachytene chromosomes of five Brassicaceae species with \( n = 5-8 \). The study corroborated results of genetic mapping, provided new insights as to the centromere positions within the ancestral karyotype, and reconstructed species-specific karyotype reshuffling of the ancestral karyotype in the five species (Lysak et al. 2006).

Another important step towards the construction of ACK was a comprehensive genetic map of the B. napus genome (Parkin et al. 2005). The authors placed 1,327 genetic loci corresponding to A. thaliana genes on the 19 linkage groups of B. napus. Knowledge about the exact position of the analyzed markers within the Arabidopsis genome made it possible to infer corresponding collinear regions within the rapeseed map. This study revealed 21 collinear blocks shared by A. thaliana and B. napus, which cover 90% of the latter genome. As the amphidiploid B. napus genome originated from an interspecific cross between B. oleracea (\( n = 9 \)) and B. rapa (\( n = 10 \)) (U 1935), structure of the two subgenomes has been reconstructed. Up to date, this is the most comprehensive comparative mapping of Brassica genomes (Schranz et al. 2006). The data obtained by Parkin et al. (2005) together with those from comparative mapping in A. lyrata and C. rubella (Boivin et al. 2004; Kuittinen et al. 2004; Yogeeswaran et al. 2005), and the reconstruction of karyotype evolution in \( n = 5-8 \) crucifers (Lysak et al. 2006) have been used to propose the model of ancestral karyotype (Schranz et al. 2006). The ACK consisting of eight ancestral chromosomes (AK1-8) and 24 conserved genomic blocks (A-X; see Schranz et al. 2006 for details) has become an important reference genome facilitating the interpretation of cross-species chromosome and genome collinearity across Brassicaceae (see also Chapter 8).

In crucifer cytogenetics, major progress in tracking evolutionary changes of the ACK has been achieved with the use of the CCP approach, whereby Arabidopsis BAC contigs are arranged correspondingly to conserved genomic blocks and then used as chromosome painting probes. Although the study of Lysak et al. (2006) can be considered as the first intentional use of the ACK in CCP, the interspecies collinearity of some genomic blocks was explored in some earlier CCP analyses (Lysak et al. 2003, 2005; Ziolkowski et al. 2006). Lysak et al. (2005) analyzed a large chromosome segment (~ 8.7 Mb in A. thaliana) corresponding to a major part of block U of the ACK in 21 species assigned to the Brassicaceae. This study allowed the time
of the polyploidization event to be estimated by comparing CCP results with the phylogeny based on the chloroplast trnL intron/trnL-F intergenic spacer region. It was shown that members of Calepina, Conringia (both genera tentatively assigned to Brassiceae), and Sisymbrium had only one copy of block U, while in the remaining Brassiceae taxa of the tribe Brassiceae three or six copies were revealed. As the genus Sisymbrium has been separated from the Brassiceae lineage and from the genus Conringia ~ 14.6 Mya (Koch et al. 2001), and the Nigra and Rapa/Oleracea clades within the Brassiceae lineage diverged ~ 7.9 Mya (Koch et al. 2001), it was concluded that the hexaploidization took place between 7.9 and 14.6 Mya (Lysak et al. 2005). Evidence for the triplication was also obtained for block F, which was studied by CCP within 10 Brassiceae species, including B. oleracea and B. rapa (Lysak et al. 2007), and block N analyzed in B. oleracea (Ziolkowski et al. 2006). Triplicated chromosomal blocks differ significantly in the gene content, as gene loss proceeds independently in each paralogous copy (Mayer et al. 2001; Rana et al. 2004). Comparison of triplicated blocks in Brassica with their homeologs in A. thaliana reveals that the gene loss is much more extensive in Brassica (Mayer et al. 2001).

Recently, CCP using Arabidopsis BAC contigs arranged according to the 24 ancestral genomic blocks of the ACK was applied to reconstruct karyotype evolution in eight x = 7 species belonging to six tribes from the phylogenetic neighborhood of the Brassiceae (Mandáková and Lysak 2008). Patterns of interspecies chromosome collinearity allowed for the reconstruction of an ancestral Proto-Calepineae Karyotype (PCK; n = 7), which was most likely shared by all x = 7 species analyzed. A detailed analysis of large-scale chromosome rearrangements was carried out, resulting in a scenario of genome evolution leading to the present-day karyotypes (Fig. 6-11). Considering the current phylogenetic relationship among the tribes and the Brassiceae (Al-Shehbaz et al. 2006; Koch and Al-Shehbaz 2009), one can conclude that the PCK karyotype could apply also to the tribe Brassiceae. Comparing the CCP patterns of the x = 7 species with the genome structure of B. napus (Parkin et al. 2005), the PCK presumably represents an ancestral karyotype of Brassiceae prior to the whole-genome triplication.

Information from ACK is also used to investigate mechanisms responsible for changes in genome structure, especially chromosome fusions, which are frequent in Brassicaceae (Lysak et al. 2006). Based on reconstruction of karyotype evolution in A. thaliana, A. lyrata, Neslia paniculata, Turritis glabra, and Hornungia alpina, a general model of chromosome number reduction was proposed (Fig. 6-12): First, a pericentric inversion converts a metacentric chromosome into an acrocentric one. Then, a reciprocal translocation occurs between the centric end of the acrocentric chromosome and a telomeric end of another chromosome. In consequence, two novel chromosomes are formed: one large “fusion” chromosome and aminichromosome. As the
minichromosome is unstable during meiosis, it can be eliminated without any significant loss of genetic information, and the chromosome number decreases (Fig. 6-12). This rearrangement could have been involved in chromosome number reduction in *Brassica* and other Brassiceae taxa, following the ancient whole-genome triplication.

Figure 6-11 Karyotype evolution in selected *x* = 7 Brassiceae tribe based on CCP (modified from Mandáková and Lysak 2008).
Up to date, the fate of the triplicated ancestral genome in the extant Brassica species was not resolved using CCP. Some data come from genetic mapping, although high level of internal genome duplication, numerous rearrangements and selective gene loss within duplicated segments hamper certain evaluation of karyotype changes. It is widely accepted that the tribe Brassiceae has monophyletic origin, thus hexaploidization took place as a result of hybridization of closely related ancestral species. Differences in rearrangement breakpoints conservation among the three copies within the B. oleracea genome (see Fig. 6-9) may suggest that the polyploidization proceeded by at least two steps, separated in time, and that Brassica hexaploid progenitor was formed by hybridization of a tetraploid genome and an ACK-like diploid genome (Ziolkowski et al. 2006; Fig. 6-13). This hypothesis needs however much more experimental research to be confirmed.

Figure 6-12 The mechanism of chromosome number reduction via a pericentric inversion and reciprocal translocation events as proposed by Lysak et al. (2006). Centromeres are white; dashed lines show breakpoints of chromosome rearrangements. Modified from Schranz et al. (2006) with permission.

Some of the important points describing the Brassica genome evolution can be drawn from cytogenetic studies. First of all, it is suggested that Brassica species separated within the family due to the whole-genome triplication (Lysak et al. 2005). As the disomic inheritance is more favorable than tetra- or hexasomic, genome diploidization has occurred. This process embraced point mutations, small-scale rearrangements as well as large chromosome aberrations (Song et al. 1995). At the chromosomal level, reciprocal translocations and inversions were the most important rearrangements, as they do not cause any genetic imbalances; these events were extensively documented by both genetic mapping and CCP studies. However, further
studies are needed to reveal chromosome rearrangements, which took place during the speciation in *Brassica*. It is likely that the extant amphidiploid *Brassica* species will also evolve through diploidization within the next few millions years, in a similar way as the hexaploid ancestor of the tribe.

![Diagram](image)

**Figure 6-13** A diagram illustrating the hypothetical genome evolution of diploid *Brassicas* via hexaploidization. Tetraploid (ancestor I) and diploid (ancestor II) ancestors are assumed, respectively. Both genomes formed a hexaploid ancestor, which gave rise to the present-day diploid *Brassicas*. The length of the arrows does not indicate evolutionary distances but only evolutionary relations (modified from Ziolkowski et al. 2006).

**Acknowledgements**

The authors wish to thank Robert Hasterok for critical reading of the manuscript. This work was partly supported by research grants from the Polish State Committee for Scientific Research (grant nos. PBZ-MNiSW-2/3/2006/19 and N/N303/313437), the Grant Agency of the Czech Academy of Science (grant no. IAA601630902) and the Czech Ministry of Education (grant no. MSM0021622415).
References

Bennett MD, Leitch IJ, Price HJ, Johnston JS (2003) Comparisons with Caenorhabditis (~100 Mb) and Drosophila (~ 175 Mb) using flow cytometry show genome size in Arabidopsis to be ~157 Mb and thus ~25 % larger than the Arabidopsis Genome Initiative estimate of ~125 Mb. Ann Bot 91: 547–557.


Cytogenetic Studies in Vegetable Brassicas 297


Lagercrantz U (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* 150: 1217–1228.


Brassica rapa Genome Sequencing Project: Strategies and Current Status

Jeong-Hwan Mun, Tae-Jin Yang, Soo-Jin Kwon, and Beom-Seok Park

ABSTRACT

Brassica rapa is considered an ideal candidate to act as a reference species for Brassica genomic studies. Among the three basic Brassica species, B. rapa (AA genome) has the smallest genome (529 Mbp), compared to B. nigra (BB genome, 632 Mbp) and B. oleracea (CC genome, 696 Mbp). There is also a large collection of available cultivars of B. rapa, as well as a broad array of B. rapa genomic resources available. Under international consensus, various genomic studies on B. rapa have been conducted, including the construction of a physical map based on 22.5X genome coverage, end sequencing of 146,000 BACs, sequencing of >150,000 expressed sequence tags, and successful phase 2 shotgun sequencing of 589 euchromatic region-tiling BACs based on comparative positioning with the Arabidopsis genome. These sequenced BACs mapped onto the B. rapa genome provide beginning points for genome sequencing of each chromosome. Applying this strategy, all of the 10 chromosomes of B. rapa have been assigned to the sequencing centers in seven countries, Korea, the UK, China, India, Canada, Australia, and Japan. The two longest chromosomes, chromosome A3 and A9, have been sequenced except for several gaps, by NAAS in Korea. The other chromosomes will be sequenced in the other six countries, over the next few years. The imminent B. rapa genome sequence offers novel insights into the...
organization and evolution of the *Brassica* genome. In parallel, the transfer of knowledge from *B. rapa* to other *Brassica* crops would be expected.

**Keywords:** *Brassica rapa*, genome, BAC, BAC-end sequence, expressed sequence tag, genetic map, physical map, chromosome sequencing

### 7.1 Introduction

The *Brassicaceae* family includes approximately 3,700 species in 350 genera. The species have diverse characteristics, many of which are of agronomic importance as vegetables, condiments, fodder, and oil crops (Beilstein et al. 2006). Economically, *Brassica* species contribute to approximately 10% of the world’s vegetable crop produce and approximately 12% of the worldwide edible oil supplies (Economic Research Service USDA 2008). The tribe Brassiceae, which is one of 13–19 tribes in the Brassicaceae, consists of ~240 species and contains most crop species of *Brassica*. Species of particular importance are *B. napus* and *B. juncea* as sources of canola oil, *B. rapa* and *B. oleracea* as vegetable cole crops, and *B. nigra* as a source of the mustard condiment. In addition to the crop species, many of the wild species in the tribe Brassiceae have potential as new crops, sources of condiments, industrial oil, and other diverse products and/or host systems for molecular farming. Wild relatives possess a number of useful agronomic traits, including nuclear and cytoplasmic male sterility, resistance to disease, insect, and nematode pests, tolerance of cold, salt, and drought stresses. For this reason, an understanding of the genetic potential of Brassiceae wild relatives is critical for the establishment of long-term breeding programs of these crops.

*Brassica* crops are characterized by diverse morphologies with enlarged vegetative and floral meristems. The morphological diversity among the *Brassica* species appears to be linked to the genomic changes associated with polyploidy (Lukens et al. 2004). Comparative genetic mapping has revealed collinear chromosome segments in the Brassicaceae family (Schmidt et al. 2001), and conserved linkage arrangements between *Arabidopsis* and *Brassica* diverged from a common ancestor approximately 14.5–20.4 million years ago (Mya) (Bowers et al. 2003). The genomes of *Brassica* species contain triplicated homoeologous counterparts of the corresponding segments of *Arabidopsis* genome, due to triplication of the entire genome that occurred approximately 13–17 Mya (O’Neill and Bancroft 2000; Town et al. 2006; Yang et al. 2006). Furthermore, an additional natural allopolyploidization event, which happened during the last 10,000 years and resulted in a change of the chromosome numbers and genome size, played a role in the diversification of *Brassica* crops (Fig. 7-1) (U 1935; Rana et al. 2004; Johnston et al. 2005).
Of the six widely cultivated *Brassica* species, *B. rapa* (AA, $2n = 20$), *B. nigra* (BB, $2n = 16$), and *B. oleracea* (CC, $2n = 18$) are the monogenomic diploids. The interspecific breeding between these three diploid species resulted in the creation of three new species of allotetraploid hybrids, namely *B. juncea* (AABB, $2n = 36$), *B. napus* (AACC, $2n = 38$), and *B. carinata* (BBCC, $2n = 34$). Thus, investigation of the *Brassica* genome provides substantial opportunities to study the divergence of gene function and genome evolution associated with polyploidy, extensive duplication, and hybridization.

![Figure 7-1 Schematic representation of the evolutionary process and genetic relationships between *Arabidopsis* and *Brassica* species with the chromosome numbers, based on literature reviewed and our research. Estimated time of major events represented as million years ago (Mya) on the left side; major events of genome divergence were summarized on the right side.](image)

The close phylogenetic relationship between the *Brassica* species and the model plant, *Arabidopsis thaliana*, implies that knowledge transfer from the results of studies on *Arabidopsis* for *Brassica* crop improvement could be straightforward. The complex genome organization of the *Brassica* species as a result of multiple rounds of polyploidy and genome hybridization, however, renders the identification of orthologous relationships of genes between the genomes highly difficult. Genome triplication, subsequent extensive interspersed gene-loss or gene-gain events, and large scale
chromosomal rearrangements including segmental duplications or deletions in the Brassica lineage all complicate the orthologous relationships of the loci between the two genomes (Town et al. 2006; Yang et al. 2006). For this reason, the genomes of several Brassica crop species have been characterized in detail over the past few years.

*B. rapa* is native to Europe and East Asia, where the existence of a large native *B. rapa* population has provided an important resource for the breeding program. There are wide morphological variations in *B. rapa*, including the leafy type (Chinese cabbage and pak choi), turnip type (vegetable turnip), and oil type (yellow sarson). Several *B. rapa* species are of regional agricultural importance, either as vegetable or oil crops. As a consequence of its native distribution and agronomic usage, *B. rapa* has great potential for use as a model for the study of both basic and applied aspects of plant biology. In particular, *B. rapa* ssp. *pekinensis* (Chinese cabbage), one of the most widely cultivated annual vegetable crops in Northeast Asia, exhibits characteristics that are useful for the study of genome characteristics such as diploidy and small genome size (529 Mbp). In response to the need for a simple genetic system with favorable genetic attributes for research among Brassica species, *B. rapa* has played a central role as a model species representing the *Brassica* “A” genome, and is the focus of multinational genome projects. Genomic studies on *B. rapa* ssp. *pekinensis* cv. Chiifu were initiated in 2003 when a bacterial artificial chromosome (BAC) library was constructed, and a collection of BAC-end sequences was initiated. Shortly afterward, a cytogenetic study, the generation of molecular markers, the construction of high-density genetic and physical maps, and eventually large-scale genome sequencing based on the clone-by-clone strategy ([http://www.brassica.info](http://www.brassica.info)) was initiated. The information and genomic resources produced during this course of investigations was useful for understanding the genetic system of *B. rapa*. Moreover, these resources will be beneficial for Brassica crop breeding, because they enable comparative genomic studies and subsequent transfer of knowledge from *B. rapa* to other Brassica crop species. In this chapter, we summarize the current state of information, the genomic resources pertaining to the *B. rapa* genome, and the progress on *B. rapa* genome sequencing, as conducted mainly by the National Academy of Agricultural Science (NAAS), RDA, Korea.

### 7.2 Characteristics of the *B. rapa* Genome

#### 7.2.1 Chromosome Structure

Cytogenetic analyses have showed that the *B. rapa* genome is organized into relatively small, compact chromosomes, with genes concentrated in the euchromatic space, unlike centromeric repeat sequences and rDNAs in
the heterochromatin which are arranged as tandem arrays (Lim et al. 2005, 2007). The individual chromosome size ranges from 2.1 µm to 4.5 µm, with a total chromosome length of 32.6 µm. An alternative cytogenetic map based on a pachytene DAPI (4',6-diamidino-2-phenylindole dihydrochloride) and fluorescent in situ hybridization (FISH) karyogram was also constructed. The mean lengths of 10 pachytene chromosomes were determined by DAPI analysis to range from 23.7 µm to 51.3 µm, with a total chromosome length of 385.3 µm. Thus, chromosomes in the meiotic prophase stage are 12 times longer than those in the mitotic metaphase, and display a well-differentiated pattern of brightly fluorescing heterochromatin segments (Koo et al. 2004).

The relative positions of 33 sequence tagged site (STS) markers and five repetitive sequences on the metaphase chromosomes were examined by FISH, and the genetic and cytogenetic correspondences between the linkage groups and chromosomes were determined (Lim et al. 2005; Kim et al. 2006). As of 2008, more than 100 gene-containing BAC clones have been analyzed by FISH, and the distribution of five major centromeric and pericentromeric repeats have been determined. By the cytogenetic method, all chromosomes can be identified based on their lengths, centromere positions, heterochromatin patterns, and positions of various repeat sequences. Centromeric satellites were estimated to encompass approximately 30% of the total chromosomes in the mitotic metaphase, particularly in the core centromeric blocks of all the chromosomes. Moreover, in the pachytene FISH, the total length of the pericentromeric heterochromatic regions was estimated to be 38.2 µm, which is approximately 10% of the total chromosome length (Koo et al. 2004). Thus, overall, heterochromatin comprises approximately 40% of the total B. rapa genome. All gene-containing BACs are localized to the euchromatin. The relationship between the cytogenetic pachytene FISH and the sequence contig distance at several locations in the euchromatin has been estimated to range from 400 to 500 kbp per µm. The FISH karyotype has created a rational basis for integrating the molecular, genetic, and cytogenetic maps of B. rapa. More importantly, it provides a basis for intelligently targeting the gene space of B. rapa and for strategic clone-by-clone genome sequencing that is likely to facilitate the discovery of most genes.

7.2.2 Repetitive Sequences and Heterochromatin

The rapidly evolving centromere structure consists of highly repetitive sequences, such as tandem satellite repeats and centromere-specific retrotransposons. The centromeric repeats characterized in plant genomes often extend over several millions of nucleotides, with 150–180 motifs such as the pAL1 satellite in Arabidopsis (Copenhaver et al. 1999), CentO in
Brassica rapa Genome Sequencing Project: Strategies and Current Status 309

rice (Zhang et al. 2004), CentC in maize (Ananiev et al. 1998), and MtRs in Medicago truncatula (Kulikova et al. 2004). The composition of the repetitive sequences in the B. rapa genome has been surveyed by a similarity search of 10,204 BAC-end sequences, with a previously reported tandem repeat that contains a HindIII site at both ends (Harrison and Heslop-Harrison 1995). Approximately 30% of the repetitive sequences showed a high similarity, and two kinds of 176 bp tandem repeats were classified with 82% sequence similarity to each other (Lim et al. 2005). From the FISH data, these 176 bp tandem repeats were localized on the centromeric regions of specific chromosomes, and named as centromeric tandem repeats of Brassica 1 & 2 (CentBr1 & CentBr2). The CentBr1 occupies centromeres on eight chromosomes (A1, A2, A4, and A6–A10) and CentBr2 resides on only 2 chromosomes (A3 and A5). This observation may reflect the predominance of CentBr1 in the B. rapa genome. The CentBr repeats were also present in other Brassica species studied, particularly in the B. oleracea whole-genome shotgun sequences, indicating that these repeats are the major components of the centromeric sequences of Brassica genome. Furthermore, sequence analysis of the heterochromatin-specific BAC clones identified additional repeat classes, including centromere-specific Ty1-copia-like retrotransposon (CRB), 238 bp-long degenerate tandem repeat (TR238) arrays, rDNAs, and pericentromere-specific Ty3-gypsy-like retrotransposons (PCRBr) (Lim et al. 2007). CRB was one of the major components of all centromeres in the three diploid Brassica species and their three reciprocal allotetraploid hybrids; however, TR238 and PCRBr were A-genome specific. Characterization of these specific centromeric or pericentromeric repeat elements may be important in identifying the heterochromatin/euchromatin borders. In addition to the centromeric or pericentromeric repeats searches, the BAC-end sequence search also identified many B. rapa-specific sequences (~ 50% of BES) that had no similarity with any sequence of Arabidopsis. This finding suggested that the B. rapa genome expansion appears to be a result of the amplification of B. rapa-specific sequences, many of which probably form heterochromatic blocks of transposons or tandem repeats (Lim et al. 2005; Yang et al. 2006).

We also identified a unique group of non-autonomous LTR retrotransposons, namely terminal-repeat retrotransposons in miniature (TRIM). TRIM elements are characterized by terminal repeats (TR), ranging from 100 to 250 bp in length, encompassing an internal domain of ~ 300 bp and creating 5 bp target site duplications. The internal sequence begins with a complement of the primer-binding site of tRNA-methionine and ends with typical polypurine tract motifs. From 96 Mbp BAC-end sequences of B. rapa, 4 distinct lineages of TRIMs (Br1–Br4), with lengths ranging from 364 bp to 1,311 bp, were identified (Yang et al. 2007). The estimated copy number of Br TRIMs was more than six times greater in the Brassica
species than in Arabidopsis, suggesting that various TRIM elements were inserted into the Brassica genome after divergence from the Arabidopsis lineage. Although the distribution of the Br TRIMs in the B. rapa genome is unknown, many of them appear to be located in the euchromatic region, based on the distribution of Arabidopsis TRIM elements in the euchromatic gene space and the occurrence of the chimeric feature of TRIMs in B. rapa expressed sequence tags (ESTs). The abundant TRIMs in the euchromatin of the B. rapa genome are expected to play an important role not only in the reconstruction of the host genome but also in the modification of the gene features by insertion of promoter or terminator sequences residing inside the elements. This modification may act as the driving force for gaining a new function, even among the duplicated genes in the Brassica genome.

The euchromatic distribution and higher insertion polymorphisms of the Br TRIMs have potential as molecular markers to distinguish the various Brassica crop species. Thus, Kwon et al. (2007) developed a transposon-display system using the unique sequences of Br1 and Br2 TRIMs. The TRIM display system successfully accessed the genetic diversity in the Brassicaceae family, and effectively identified 16 commercial F1 hybrids of B. rapa and other Brassica crops (Kwon et al. 2007).

7.2.3 Triplicated Nature of the B. rapa Genome

One of the important challenges in plant genomics is to redirect the knowledge gained from the study of model genomes into the important biological and agronomical questions among crop species, based on whole-genome sequencing of model genomes. Comparative genomic studies are well established in several plant families such as cereals and legumes (Paterson et al. 2003; Choi et al. 2004; Cannon et al. 2006). Because Brassica species are closely related to A. thaliana, comparative genomic studies are aimed to elucidate the Brassica genome organization and evolution for comparison with the A. thaliana genome. One of the profound findings from the comparative analysis was the recognition of the triplicated nature of the Brassica genome, indicating whole-genome triplication after the split from the A. thaliana lineage (Lysak et al. 2005; Town et al. 2006; Yang et al. 2006).

The triplicated nature of the B. rapa genome was revealed by comparative genetic and physical mapping and sequence level synteny comparison. The foundation of the comparative mapping has been a restriction fragment length polymorphism (RFLP) marker-based genetic map of cv. Jangwon. Kim et al. (2006) used 545 genetic markers based on B. rapa ESTs and Arabidopsis genes. Among the 520 RFLP markers, a majority of the markers detected > 1 locus, and only 62 EST markers showed a single locus anchor (Kim et al. 2006). Interchromosomal comparison of the linkage groups identified
many homologous blocks in the genome. Comparative physical mapping between *A. thaliana* and *B. rapa* further revealed triplicated homologous blocks, including FLOWERING LOCUS C (FLC) regions in the *B. rapa* genome with collinearity of genes in the blocks (Park et al. 2005). These findings collectively suggested that the diploid *B. rapa* genome has been derived from a hexaploid ancestor. Sequence level analysis of the triplicated FLC regions of *B. rapa* revealed the diploidization process in the triplicated genome. Yang et al. (2006) compared four paralogous bacterial artificial chromosome (BAC) clones covering the FLC loci and the homologous 124 kbp segment of *Arabidopsis* chromosome 5. They estimated the time elapsed since the divergence of the paralogous and homologous lineages, and reported that the three paralogous subgenomes of *B. rapa* triplicated around 13–17 MYA, soon after the *Arabidopsis* and *Brassica* divergence occurred around 17–18 Mya (this result was summarized in Fig. 7-1). In addition, the *B. rapa* genome underwent segmental duplication around 0.8 Mya. The *B. rapa* genome segments showed extensive interspersed gene loss and insertion of specific transposons such as TRIMs relative to the *Arabidopsis* genome segment. The diploidization process was estimated to reduce 40% of the triplicated genome length, meaning that the *B. rapa* genome would contain only approximately a 1.7-fold higher number of genes than *Arabidopsis*.

The duplicated genes retained in the *B. rapa* genome may have a selective advantage owing to the gain of new functions (neofunctionalization) or partitioning of the original function (subfunctionalization) between the two duplicates. Genome triplication and diploidization in the *Brassica* lineage would affect the functional diversification of redundant genes, because these events could permit mutations in the ancestor genome loci that are normally under tight selective constraints; the resulting subfunctionalization or neofunctionalization of the duplicated genes might lead to phenotypic diversification of *Brassica*. A study on the functional diversification of triplicated genes would provide an insight into the role of polyploidization in the *Brassica* crop genomes.

**7.3 Genomic Resources for *B. rapa***

**7.3.1 BAC Libraries and BAC-end Sequences**

Various genomic resources are indispensable for genomic study of any crop species. The genomic resources available for the *B. rapa* Genome Sequencing Project (BrGSP) are summarized in Table 7-1. A successful structural genomic study of the *B. rapa* genome relies on the quality and availability of detailed large-insert genomic libraries. As of 2008, five large-insert BAC libraries of *B. rapa* ssp. *pekinensis* cv. *Chiifu* were publicly available, providing approximately 53-fold genome coverage overall. These libraries were constructed using the restriction enzymes *Eco*RI, *Bam*HI, *Hind*III, and *Sau*3AI.
<table>
<thead>
<tr>
<th>Resources</th>
<th>Material</th>
<th>Characteristics</th>
<th>Number*</th>
<th>Referencea</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAC library</td>
<td>5 libraries</td>
<td></td>
<td>234,544 clones (53.4X)</td>
<td>KBGP</td>
</tr>
<tr>
<td>KBrB</td>
<td>cv. Chiifu / BamHI</td>
<td>avr. insert size 120 kbp</td>
<td>55,296 clones (12.5X)</td>
<td>KBGP, NCBI</td>
</tr>
<tr>
<td>KBrE</td>
<td>cv. Chiifu / EcoRI</td>
<td>avr. insert size 139 kbp</td>
<td>23,040 clones (6X)</td>
<td>KBGP, NCBI</td>
</tr>
<tr>
<td>KBrH</td>
<td>cv. Chiifu / HindIII</td>
<td>avr. insert size 120 kbp</td>
<td>56,448 clones (12.8X)</td>
<td>KBGP, NCBI</td>
</tr>
<tr>
<td>KBrS1</td>
<td>cv. Chiifu / Sau3AI</td>
<td>avr. insert size 100 kbp</td>
<td>55,296 clones (10.5X)</td>
<td>KBGP, NCBI</td>
</tr>
<tr>
<td>KBrS2</td>
<td>cv. Chiifu / Sau3AI</td>
<td>avr. insert size 123 kbp</td>
<td>46,464 clones (11.6X)</td>
<td>KBGP, NCBI</td>
</tr>
<tr>
<td>BAC-end sequence</td>
<td>146,688 clones</td>
<td>both end</td>
<td>260,637 reads (203 Mbp)</td>
<td>KBGP, NCBI</td>
</tr>
<tr>
<td>KBrB</td>
<td>55,296 clones</td>
<td>single-pass sequence</td>
<td>97,912 reads (75 Mbp)</td>
<td>KBGP, NCBI</td>
</tr>
<tr>
<td>KBrE</td>
<td>23,040 clones</td>
<td></td>
<td>43,168 reads (33 Mbp)</td>
<td>KBGP</td>
</tr>
<tr>
<td>KBrH</td>
<td>50,688 clones</td>
<td></td>
<td>88,951 reads (73 Mbp)</td>
<td>KBGP, NCBI</td>
</tr>
<tr>
<td>KBrS1</td>
<td>6,144 clones</td>
<td></td>
<td>8,117 reads (5 Mbp)</td>
<td>KBGP, NCBI</td>
</tr>
<tr>
<td>KBrS2</td>
<td>11,520 clones</td>
<td></td>
<td>22,489 reads (17 Mbp)</td>
<td>KBGP</td>
</tr>
<tr>
<td>Genetic map</td>
<td>3 reference maps</td>
<td></td>
<td>1,287 cM, 545 markers</td>
<td>Kim et al. 2006</td>
</tr>
<tr>
<td>JWF3p</td>
<td>cv. Jangwon</td>
<td>F$_2$ population 134 lines</td>
<td>1,182 cM, 556 markers</td>
<td>Choi et al. 2007</td>
</tr>
<tr>
<td>CK</td>
<td>cv. Chiifu × cv. Kenshin</td>
<td>DH F$_2$ population 78 lines</td>
<td>KBGP</td>
<td></td>
</tr>
<tr>
<td>VCS</td>
<td>cv. VC1 × cv. SR5</td>
<td>DH F$_2$ population 91 lines</td>
<td>KBGP</td>
<td></td>
</tr>
<tr>
<td>Physical map</td>
<td>Build 2</td>
<td>67,468 BAC fingerprints</td>
<td>1,428 contigs (717 Mbp)</td>
<td>Mun et al. 2008</td>
</tr>
<tr>
<td>Expressed sequence</td>
<td>33 cDNA libraries of cv. Chiifu</td>
<td>cDNA single-pass sequence</td>
<td>KBGP</td>
<td></td>
</tr>
<tr>
<td>Microarray</td>
<td>2 NimbleGen chips</td>
<td></td>
<td>152,253 ESTs (91 Mbp)</td>
<td>KBGP</td>
</tr>
<tr>
<td>KBGP-24K</td>
<td>24,000 unigenes</td>
<td>60mer, 6 probes/gene</td>
<td>Lee et al. 2008</td>
<td></td>
</tr>
<tr>
<td>KBGP-50K</td>
<td>32,000 unigenes, 17,000 CDS</td>
<td>60mer, 7 probes/gene</td>
<td>KBGP</td>
<td></td>
</tr>
<tr>
<td>BAC sequence</td>
<td>KBrB, KBrE, KBrH, KBrS clones</td>
<td>BAC shotgun sequence</td>
<td>Ongoingc</td>
<td>KBGP, NCBI</td>
</tr>
</tbody>
</table>

*Total sequence length or genome coverage is represented in parenthesis. Genome coverage was estimated based on the haploid genome equivalent of B. rapa as 529 Mbp.

aKBGP, Korea B. rapa Genome Project (http://www.brassica-rapa.org), NCBI, the National Center for Biotechnology Information.

cAs of December 2008, 886 BAC sequences have been deposited in NCBI.
Using these BAC libraries, a total of 260,637 BAC-end sequences (BES) have been generated from 146,688 BAC clones (~203 Mbp), as a collaborative outcome of the multinational BrGSP community. Analysis of BES, combined with BAC sequence surveys, enabled the outlining of the features of the whole-genome structure of *B. rapa*. BLAST search of BES identified that up to 25% of BES were estimated to contain centromeric or pericentromeric repetitive sequences. An additional 15% of BES were matched with transposons and other repeat sequences. Based on this data and FISH analysis (see Section 7.2.1), the heterochromatic region was postulated to occupy >40% of the *B. rapa* genome, while the euchromatic gene space was postulated to constitute <60% of the *B. rapa* genome (Yang et al. 2005; Lim et al. 2007). Comparison of the BES with the *B. rapa* ESTs and *Arabidopsis* CDS led to the recognition of approximately 11% of the BES as protein-coding genes. Assuming that the average CDS size of the BAC survey sequences ranges from 1.1 to 1.3 kbp, the estimated number of genes in the whole genome would be at least 45,000 to 53,000, which is roughly similar to the estimation of Yang et al. (2006).

Comparison of the BES with the *Arabidopsis* genome was an efficient application of the BES dataset to select the seed BAC clones for *Brassica* genome sequencing (Yang et al. 2005). In silico comparative sequence matching of 91,000 *B. rapa* BES to the *A. thaliana* chromosome sequences identified approximately 50% of the BES showing significant sequence similarity along with overall collinearity to counterpart *Arabidopsis* chromosomal regions. Based on the comparative genetic map and microcollinearity between the two genomes, the BAC clones mapped onto the *Arabidopsis* genome can be chosen as seed BAC clones, even before a complete physical map is established. Practically, approximately 600 seed BAC clones were selected and sequenced by this method. The details of BES matching and seed BAC clone selection are explained in the section on genome sequencing (see Section 7.4.2). Almost all the BAC-end sequenced clones were also fingerprinted, and the BES data and physical contig information was combined in the BAC extension program of genome sequencing to select the BAC clones from a seed point to be sequenced. This tool uses the BLAST algorithm to match one or more input sequences with any sequence associated with the fingerprinted clones, similar to the Blast Some Sequence (BSS) tool in the FingerPrinted Contig (FPC) program.

### 7.3.2 Genetic Map

A genetic map is the basic starting point for any genomic study or sequencing. Genetic maps of *B. rapa* have been reported using various mapping populations. Most of the maps were constructed using hybridization-based markers. Recently, three high-density genetic maps of *B. rapa* ssp. *pekinensis*
have been reported as references to guide the ongoing genome sequencing. One of them is an intercultivar map based on a cv. Chiifu × cv. Kenshin cross (Choi et al. 2007). The CK map used 78 doubled haploid lines, providing direct and accurate genetic information for Chiifu genome sequencing. For this reason, the BrGSP community chose the CK map as the reference map for the ongoing genome sequencing project. The CK map consists of a total of 556 markers over a length of 1,182 cM; however, the number of anchored BACs in the sequencing pipeline is limited. The second map is based on an intracultivar $F_{2:3}$ linkage map based on the cultivar Jangwon cross (Kim et al. 2006), and was first developed using 520 RFLP markers and 25 PCR-based markers over a length of 1,287 cM. Although the Jangwon map does not represent perfect information for Chiifu sequencing, it plays a crucial role in determining the positioning of the sequenced BAC clones. Most of the RFLP markers of the Jangwon map have been replaced by PCR-based sequence tagged site (STS) markers, including gene-targeting or simple sequence repeat (SSR) markers designed on the BAC sequences. As of 2008, a total of 513 sequenced BACs have been anchored on the Jangwon map, which contains 905 genetic loci over a length of 1,311 cM. A third genetic map has been established from a cross between cv. VC1 and cv. SR5 (http://www.brassica-rapa.org). The VCS map gives the positions of 225 sequenced BACs that are not included in the Jangwon map. In addition, the VCS map provides enhanced genetic resolution in regions where suppression of recombination or distorted segregation was shown in the previous maps. Integration of these three maps is underway using a set of common anchored markers, and an integrated map will soon be available to the sequencing community.

The availability of molecular markers, high-density genetic maps, and the genome sequence has significantly accelerated the identification and isolation of genes that are important for crop quality improvement and development, for example, the genes pertaining to disease resistance, flowering time, or biosynthesis of valuable phytochemicals. In addition, high levels of microsynteny between B. rapa and other Brassica species allows for the possibility of using syntenic positional cloning as a basis for isolating orthologous genes of interest in other crops of the Brassicaceae family.

7.3.3 Physical Map

The availability of a genome-wide, sequence-ready physical map is one of the crucial components for a successful clone-by-clone strategy. A physical map not only makes it possible to determine clones for genome sequencing with comprehensive coverage and reduced sequencing redundancy, but also enables one to simplify the sequence assembly by arranging the sequence contigs in order. Thus far, the utility of physical
maps has been reported by major genome sequencing projects on the human (International Human Genome Sequencing Consortium 2001), *A. thaliana* (Marra et al. 1999), rice (Chen et al. 2002), and *M. truncatula* (Mun et al. 2006). These physical maps were constructed using a combination of techniques, including restriction enzyme-digested BAC fragment fingerprinting on agarose gels and assembly of the fingerprints using the FPC software package (Soderlund et al. 2000). The agarose method has been successful, but has limited throughput because of the need for manual band calling. Alternatively, fluorescence-labeled fingerprinting methods have been developed to make larger and more accurate contigs, with increased throughput using an automatic capillary sequencer (Gregory et al. 1997; Ding et al. 2001; Luo et al. 2003; Xu et al. 2004). Fluorescence-labeled capillary electrophoresis methods include the 3-enzyme method and the high-information content fingerprinting (HICF) methods, which use type IIS restriction enzymes or SNaPshot labeling techniques, respectively. As compared to the agarose method, the automatic workflow and higher resolution of these methods facilitates better physical map construction, both in terms of throughput and the quality of fingerprinting. The first genome-wide plant HICF physical map was constructed for maize (Nelson et al. 2005).

We constructed a genome-wide BAC-based physical map of *B. rapa* by the SNaPshot method (Mun et al. 2008). To create a robust sequence-ready physical map, a total of 99,456 BAC clones from the three independent BAC libraries (~ 22.5 × coverage) were fingerprinted by digestion with combinations of five restriction enzymes (*Eco*RI, *Bam*HI, *Xba*I, *Xho*I, and *Hae*III), followed by SNaPshot reagent labeling of four colors at the 3’-ends of the restriction fragments, then sizing on the ABI 3730x1 capillary sequencer. Of the fingerprints, a total of 93,689 clones (94.2%) were successfully fingerprinted to be used for contig assembly. From the initial dataset, 26,221 BAC clones containing heterochromatic repetitive sequences were removed from the contig assembly, which significantly enriches the euchromatic contigs in the resulting build. The physical contig map was assembled using 67,468 high-quality, heterochromatic repeat-free BAC fingerprints from the initial dataset. These BAC clones represent 15.2× coverage of the *B. rapa* genome; they were condensed into 1,417 contigs, and the resulting contigs were manually edited to validate reliability. With the results of the contig evaluation, manual editing of the initial contig build yielded 1,428 contigs, with an average length of 512 kbp spanning 717 Mbp, 1.3× coverage of the genome. An unsatisfactory aspect of this assembly is its large number of Q clones, as the Q clones in this assembly corresponded to 15% of the clones. However, 3 specific deep contigs contributed to ~ 48.3% of all the Q clones in the build. Thus, when these deep contigs of the initial build were excluded due to false-positive overlaps, the Q clones in the remaining contigs correspond to 7.7% of all clones. This ratio is similar to the ratios
reported in catfish (7.3%) (Quiniou et al. 2007) and maize (11%) HICF maps (Nelson et al. 2005).

The contigs produced in the course of the fingerprinting work were tagged with 315 anchored genetic markers. Practically, the important aspect of this was the integration of a physical map into a genetic map, enabling the positioning of 242 gene-rich contigs to specific locations on 10 chromosomes and providing seeds for the genome sequencing effort. The number of contigs associated with genetic loci is ~160.7 Mbp, or 30% of the total genome. The total coverage of the physical contigs suggests that most contigs do not have sufficient overlaps, and the gaps between the contigs need to be filled by additional fingerprinting. To improve the map, additional fingerprinting of approximately 30,000 clones of 2 new BAC libraries (KBrE and KBrS2) was performed. This data is being merged into the current build to continue the refinement of the physical map. In 2009, updating of the physical map was kept open for the B. rapa genome sequencing consortium. In parallel, linkage analysis of SSRs and single nucleotide polymorphisms (SNPs) in the seed BACs and BESs will be carried out to provide more anchoring points for the physical map. The details of these BACs and the associated markers can be found on the Korea B. rapa Genome Project (KBGP) website of NAAS (http://www.brassica-rapa.org). This information plays an important role in identifying the BAC clones for sequencing.

7.3.4 Expressed Sequence Tags and Transcriptome Analysis

To support genome annotation and functional study, we have sequenced a total of 152,253 B. rapa ESTs. These ESTs are obtained from 33 cDNA libraries representing a variety of organs and development stages and have been deposited in public databases. These ESTs were clustered into 39,095 unique sequences (unigenes), including 16,898 tentative consensus sequences. The current collection of B. rapa ESTs consistently represents the whole B. rapa genome, with the Arabidopsis genome as a reference. In silico mapping of the 39,095 unigenes on the Arabidopsis genome identified 85% of the unigenes, covering 75% of the overall Arabidopsis counterpart coding sequences. It was found that the remaining 15% of the B. rapa unigenes were not homologous with any gene in A. thaliana, rather, this 15% represented novel B. rapa-specific genes. Gene ontology analysis of the ESTs did not show any significant overestimation of specific categories in the B. rapa genome, as compared to the Arabidopsis genome. Analysis of the EST collections identified 21,409 full-length cDNA sequences. It is anticipated that these sequences will contribute to the formation of a B. rapa-specific set of sequences, not only for gene prediction programs for B. rapa and other Brassica species, but also to evaluate structure and alternative splicing of the predicted gene models.
The corresponding cDNA clones have been used to construct microarrays for expression profiling during development, under various biotic- and abiotic-stress conditions. Two microarrays, namely KBGP-24K and KBGP-50K, were developed using the NimbleGen platform. Both microarrays included the six (KBGP-24K) or seven (KBGP-50K) 60-nucleotide-long probes per gene. The 24K chip covered approximately 24,000 unigenes clustered by 127,144 ESTs from 20 cDNA libraries, whereas the 50K chip doubled the gene contents by including an additional 8,500 unigenes, plus 17,500 genes predicted from the seed BAC sequences in the genome sequencing pipeline. These microarrays examined the changes in the genome-wide gene expression of *B. rapa*, in response to transcriptional changes. Using the KBGP-24K chip, genome-wide transcriptome analysis was conducted in response to three abiotic stresses that significantly affect the productivity of *Brassica* crops: salt, cold, and drought (Lee et al. 2008). This analysis successfully identified stress-related genes along with novel transcription factor genes, suggesting the existence of a *B. rapa*-specific signaling pathway that works together with the common stress-response pathway under abiotic stress conditions. Information on the EST, unigene, and microarray experiments can be retrieved from the *B. rapa* EST and Microarray Database (BrEMD) website (http://www.brassica-rapa.org/BrEMD).

### 7.3.5 Functional Genomics and Information Resources

The online database (http://www.brassica-rapa.org) provides easy access to genomics information such as nucleotide sequences of the BAC clones with tentative annotation of the predicted genes, genetic maps along with the molecular markers, a physical map that can be viewed on WebFPC, cDNA sequences, and various transcriptome profiling data obtained by KBGP-24K microarray analysis. The BAC and cDNA libraries and clones can also be ordered through this site. The Korea *Brassica* Genome Resource Bank (KBGRB, http://www.brassica-resource.org) is a material resource database of *B. rapa*, and is supported by the Korea National Plant Research Resource Center. Resources can be obtained through this database, including the seeds of various *B. rapa* wild accessions and cultivars, recombinant inbred (RI) lines, mapping populations, and BAC and cDNA libraries.

A large number of T-DNA insertion mutant lines are being generated as resources for reverse genetics, and a targeting induced local lesions in genomes (TILLING) system initiated by the Multinational *Brassica* Genome Project (MBGP) TILLING Consortium has been developed. TILLING is a high-throughput, reverse genetic approach that combines the efficiency of
ethylmethane sulfonate (EMS) with high-throughput genotyping methods to detect SNPs. The MBGP TILLING Consortium aims to share the EMS mutant populations developed from the Chiifu (Korea), rapid-cycling (USA), R-O-18 (UK), and L58 (China) cultivars for functional genomic studies. The large quantity of genetic information provided by the BrGSP will make the TILLING approach efficient and feasible.

7.4 Progress and Current Status of B. rapa Genome Sequencing

7.4.1 Sequencing of Euchromatic Regions Based on the Clone-by-Clone Strategy

Despite the importance of Brassica crops in plant biology and world agriculture, none of the Brassica species have been fully sequenced. Considering the importance of the Brassica species and the value of B. rapa as a model Brassica, sequencing most of its gene spaces will be very beneficial. Clone-by-clone sequencing and whole-genome shotgun (WGS) sequencing are both well-established genome sequencing approaches, and can be used to investigate the whole-genome structure of the species. We selected the clone-by-clone genome sequencing strategy because of its accuracy for the complex triplicated nature of the Brassica genome. The clone-by-clone strategy typically involves a “minimum tiling path” of large-insert (~100 kbp) clones, such as BACs of a known order, which is determined using a combination of genetic, physical, and/or cytogenetic mapping. Cytogenetic studies based on extensive fluorescence in situ hybridization (FISH) analysis of both metaphase and pachytene chromosomes have provided detailed insight into the organization of heterochromatic and euchromatic regions. These cytogenetic studies demonstrated that the genome of B. rapa is organized into distinct repetitive sequence-rich pericentromeric heterochromatin and gene-rich euchromatin. Sequencing of several BAC clones has confirmed that the gene density in B. rapa is relatively high on the order of one gene per 3–4 kbp (Yang et al. 2006). Each of the gene-rich BAC clones examined so far by FISH (>100 BACs) was found to be localized to the visible euchromatic region of the genome. Concurrently, a whole-genome shotgun pilot sequencing of B. oleracea with 0.44-fold genome coverage generated transposable element-enriched sequences (Zhang and Wessler 2004; Ayele et al. 2005). Taken together, these data strongly point to a favorable genome organization where the overwhelming majority of the B. rapa euchromatic space can be sequenced in a highly efficient manner by the clone-by-clone strategy.
7.4.2 Initial Approach for Selection of Seed BACs Based on Comparative Genomics

The clone-by-clone sequencing starts from the defined seed points and builds outward. The fingerprint-based physical map, combined with BES and genetic anchoring data, provides a basis for selecting seed BAC clones and for creating a draft tiling path. Alternatively, comparative approaches using an already sequenced closely related model genome, such as the comparative tiling method, can be used as a backbone for in silico clone validation of seed BACs, even before the availability of a physical map. Indepth comparative sequence analyses of several sequenced \textit{B. rapa} BAC clones revealed overall collinearity with a homologous region of the \textit{Arabidopsis} counterpart. Thus, if we allocate \textit{B. rapa} BAC clones on the \textit{Arabidopsis} euchromatin using BES matches and scatter them onto the \textit{B. rapa} chromosomes, selection of gene-rich BAC clones will be straightforward even without the information from the physical map (Fig. 7-2). Based on this idea, Yang et al. (2005) compared the BES of 46,000 \textit{B. rapa} BAC clones with \textit{Arabidopsis} sequences using BLASTZ. They mapped the BAC clones on the five \textit{Arabidopsis} chromosomes by in silico matching based on unique, significant (< E$^{-6}$), and directional matches of the pair ends sequences of each BAC with a complement match within the 30–500 kbp interval (Yang et al. 2005). Approximately 50% of the BES showed significant sequence similarity with unique \textit{A. thaliana} sequences, and a total of 4,317 BAC clones were mapped onto the \textit{Arabidopsis} chromosomes by significant matching with both ends within the 30-500 kbp interval. These \textit{B. rapa} BAC clones spanned 93 Mbp of the \textit{A. thaliana} sequences, representing ~ 78% of the total \textit{Arabidopsis} genome. BAC-FISH and STS mapping using BES of selected BAC clones positioned on the counterpart \textit{Arabidopsis} chromosomes showed the real euchromatic locations of the BAC clones scattered on the \textit{B. rapa} chromosomes. A single \textit{B. rapa} BAC clone was calculated to span an average of 147 kbp of the \textit{A. thaliana} counterpart sequence. Theoretically, 500 contiguous BAC clones will cover around 80 Mbp of the euchromatic regions of the \textit{Arabidopsis} genome, assuming the average insert size of a BAC clone is 120 kbp. Therefore, if minimally overlapping BAC clones mapped onto the \textit{Arabidopsis} genomes were selected and scattered onto the \textit{B. rapa} chromosomes, they could provide a seed point for bidirectional outward genome sequencing. As of July, 2008, 589 minimally tiled \textit{B. rapa} BAC clones spanning 75 Mbp of the \textit{A. thaliana} genome were sequenced in phase 3 (finished sequences) or phase 2 (sequences that are fully orientated and ordered, but contain some small sequence gaps and low-quality regions). Most of them were distributed onto the 10 \textit{B. rapa} chromosomes by STS mapping, FISH analysis, and physical contig information (Fig. 7-2). The 589 sequenced BAC clones were provided to BrGSP as seed BACs and used as starting points for chromosome
sequencing (data available on the KBGP website). Integration of seed BACs with the physical map provides “gene-rich” contigs spanning ~160 Mbp (see Section 7.3.3). These “gene-rich” contigs will enable us to select clones to extend the existing sequence contigs.

Figure 7-2 Schematic representation of the comparative tiling sequencing method to select seed BAC clones from the euchromatic regions of *B. rapa*, using in silico matching of BESs with known *Arabidopsis* sequences. (A) Based on the knowledge of the sequence level collinearity between *Arabidopsis* and *Brassica*, BAC-end sequences of *B. rapa* ssp. *pekinesis* cv. *Chiifu* were BLAST analyzed against *Arabidopsis* genome sequence and the corresponding BACs were selected based on the pair-end best hits within 30–500 kbp interval. (B) Selection of minimally overlapping 589 BAC clones based on in silico comparative allocation on the *Arabidopsis* genome. A total of 4,317 *B. rapa* BAC clones with an average insert size of 120 kbp were mapped in silico onto the euchromatic regions of the *Arabidopsis* genome. On in silico comparative allocation, 589 minimally overlapping BAC clones were mapped onto the comparative positions in the corresponding *Arabidopsis* genome sequence, then were selected and sequenced as phase 2 and scattered on the euchromatic regions of the 10 *B. rapa* chromosomes as seed BACs by a combination of genetic and cytogenetic mapping. Mixed-color bar represents each chromosome of *Arabidopsis* and blue bars depict *B. rapa* BAC contigs tiled on the counterparts of *Arabidopsis* genome. (C) Based on the sequence of each BAC clone, we developed STS markers and anchored the clones on the genetic maps. Overall, 513 sequenced BAC clones were mapped onto the 10 linkage groups and provided to each sequencing center of the BrGSP for completion of the chromosome sequencing. Detailed numbers of BAC clones on each chromosome are denoted in Table 7.3.

*Color image of this figure appears in the color plate section at the end of the book.*
7.4.3 Features of Protein Coding Genes

The predicted structures of the potential protein coding genes in the seed BAC clones, deposited in the High Throughput Genome Sequence (HTGS) database of the National Center for Biotechnology Information (NCBI), have been predicted ab initio using FGENESH (www.softberry.com) based on a B. rapa matrix (Table 7-2). Although the structural features of the protein coding genes in B. rapa are similar to those of the B. oleracea genes reported previously (Town et al. 2006), the former show a smaller average gene length (1.6 kbp) as compared to Arabidopsis (2.2 kbp). This difference appears to amount to almost one less exon per gene (4.7 and 5.8 exons per gene in B. rapa and A. thaliana, respectively) along with a shorter exon (225 bp in B. rapa and 268 bp in A. thaliana, respectively) and intron length (141 bp in B. rapa and 165 bp in A. thaliana, respectively) in B. rapa. The gene density in the sequenced BAC clones of the B. rapa genome (one gene per 4.2 kbp) is higher than that in the Arabidopsis genome (one gene per 4.5 kbp), indicating the compact organization of the euchromatic region of B. rapa. A similarity search of the protein coding genes of B. rapa against public databases indicated that approximately 18% of the predicted genes showed no significant similarity with any of the genes reported. This result is roughly consistent with the results of previous studies conducted using EST analysis (see Section 7.3.4) and synteny analysis of FLC regions (Yang et al. 2006). NAAS, the leading sequencing center for BrGSP, is developing an automated genome annotation pipeline that combines gene prediction and homology data, such as B. rapa ESTs and full-length cDNAs, to produce the best working model. High-quality sequences and first-pass annotation will be released by the center to the user community in a timely fashion.

Table 7-2 Comparison of overall composition of annotated protein coding genes in the B. rapa BAC sequences and A. thaliana genome.

<table>
<thead>
<tr>
<th>Feature</th>
<th>B. rapa</th>
<th>A. thaliana*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average gene size (kbp)</td>
<td>1.6</td>
<td>2.2</td>
</tr>
<tr>
<td>Number of exons per gene</td>
<td>4.7</td>
<td>5.8</td>
</tr>
<tr>
<td>Intron size (bp)</td>
<td>141</td>
<td>165</td>
</tr>
<tr>
<td>Exon size (bp)</td>
<td>225</td>
<td>268</td>
</tr>
<tr>
<td>Average gene density (kbp/gene)</td>
<td>4.2</td>
<td>4.5</td>
</tr>
<tr>
<td>Overall G/C content (%)</td>
<td>35.2</td>
<td>35.9</td>
</tr>
<tr>
<td>Exons</td>
<td>46.3</td>
<td>44.1</td>
</tr>
<tr>
<td>Introns</td>
<td>32.6</td>
<td>32.6</td>
</tr>
<tr>
<td>Intergenic regions</td>
<td>31.3</td>
<td>32.9</td>
</tr>
</tbody>
</table>

*Statistics for A. thaliana is based on the annotation of TAIR7 from the Arabidopsis Information Resource website (http://www.arabidopsis.org/portals/genAnnotation/genome_snapshot.jsp).
7.4.4 Sequencing Status and Progress

The BrGSP has decided to sequence the euchromatic arms of all 10 chromosomes, with sequencing responsibilities distributed as represented in Fig. 7-3. The project aims to initially produce a “phase 2” sequence with accessible trace files by shotgun, so that the researchers who require finished sequences from a specific region can finish them. BAC clones are sequenced by shotgun and standard assembly methods, then gene annotation is conducted by a combination of a semiautomatic method and manual editing. STS markers (including SSR or SNP markers generated from the BAC sequences) will localize BAC clones onto the reference genetic map.

![Figure 7-3 Idiogram of B. rapa chromosomes represented by euchromatic arms and heterochromatic repetitive sequence blocks. Each chromosome was assigned to a sequencing center in one of the seven countries of the BrGSP consortium.](image)

Although the details are different, each sequencing center has adopted similar BAC sequencing strategies. Large-insert BAC clones are isolated by a semi-automated “cleared lysate” procedure and are tested for clone fidelity. Shotgun libraries are produced from both small and large inserts generated by hydrodynamic shear, followed by size selection and cloning. The resulting size-fractionated, sheared fragments are blunt-end cloned after end-filling and agarose gel sizing. The libraries are sequenced, up to a total of 7- to 8-fold sequence coverage. The sequencing reactions are optimized to maximize the read lengths on the ABI 3730xl sequencers. Generally, 700-bp high-quality sequences are obtained by capillary running
and read after vector trimming. Base-calling and sequence assembly are performed using the Phred/Phrap/Consed software package, and the new assemblies are released to the HTGS database of NCBI. Sequencing technology is a dynamic field, and it is possible that opportunities that will enable the centers to sequence more efficiently may arise during the course of the project. Advances in sequencing technology using next-generation sequencing machines could lower the sequencing costs. In fact, NAAS tested a combination of GS-FLX and ABI 3730xl sequencing machines that could reduce the sequencing costs by almost 20%; however, the lower resolution of the poly-mononucleotide tracks in the pyrosequencing data required additional validation.

As of December 2008, a total of 107 Mbp from 886 BAC clones (401 phase 3, 445 phase 2, and 40 phase 1 clones) have been sequenced and deposited in the HTGS database of NCBI by 3 sequencing centers: Korea (681 BACs), UK/China (200 BACs), and Australia (five BACs). The other participating sequencing center will soon increase its capacity to facilitate sequencing. Besides the reported BAC clones, NAAS sequenced an additional 884 BAC clones as a result of sequencing of the chromosomes A3 and A9 (Table 7-3). Using 670 BAC clones, the sequence scaffolds for the two chromosomes were constructed. In the case of A3, 381 minimum tiled BAC sequences generated seven sequence scaffolds spanning 34.9 Mbp; however, in the case of A9, 289 minimally tiled BAC clones generated 15 scaffolds comprising 28.5 Mbp (Fig. 7-4). Based on the estimated size of the \textit{B. rapa} euchromatic region, the assemblies thus far constitute 90\% and 85\% of the entire euchromatic region of chromosome A3 and A9, respectively (our unpublished data). Additional effort is ongoing, in an attempt to fill the gaps between the scaffolds. The details of the sequences will be released soon with annotation.

Table 7-3: Current status of \textit{B. rapa} BAC sequencing by NAAS, as of March 2009.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>A8</th>
<th>A9</th>
<th>A10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchored seed BACs</td>
<td>56</td>
<td>62</td>
<td>94</td>
<td>14</td>
<td>53</td>
<td>58</td>
<td>50</td>
<td>47</td>
<td>58</td>
<td>34</td>
</tr>
<tr>
<td>Extended BACs</td>
<td>1</td>
<td>5</td>
<td>441</td>
<td>6</td>
<td>1</td>
<td>14</td>
<td>7</td>
<td>7</td>
<td>291</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>57</td>
<td>67</td>
<td>535</td>
<td>20</td>
<td>54</td>
<td>72</td>
<td>57</td>
<td>54</td>
<td>349</td>
<td>46</td>
</tr>
</tbody>
</table>

### 7.5 Conclusion

\textit{B. rapa} will provide an excellent reference genome sequence, as well as valuable information for understanding the genetic systems of the \textit{Brassica} crop species. The most significant beneficiaries of \textit{B. rapa} sequences will be \textit{Brassica} crop researchers, from breeders to plant biologists. A large number of \textit{B. rapa} EST collections, full-length cDNA sets, and the genome sequences will enable the identification and isolation of the genes of interest for agriculture. High-density genetic maps, STS markers, and associated sequence information have accelerated genome mapping and map-based
cloning. The comparative genomics approach for *B. rapa* will also benefit genomic investigation of closely related *Brassica* crops, including *B. oleracea* and *B. napus*. Quantitative trait loci (QTL) or association mapping of valuable phytochemical-related genes in *B. napus* utilizes the available *B. rapa* genome sequence data and resources (Meng, Huazhung, pers. comm.). In addition, genome sequencing of other *Brassica* crops, particularly the construction of sequence assemblies and scaffolds of *B. napus*, will be improved owing to the presence of information on the *B. rapa* genome. Technically, newly emerging next generation sequencing techniques using GS FLX Titanium (Roche), Genome Analyzer (Illumina), and SOLiD (ABI) systems can be

Figure 7-4 Current sequencing status of the two longest chromosomes, A3 and A9. BAC clones were sequenced as phase 2 and assembled into sequence scaffolds. (A) A total of 381 minimally tiled BAC clones generated seven sequence scaffolds spanning 34.9 Mbp for chromosome A3. (B) 15 scaffolds spanning 28.5 Mbp were assembled from 289 minimum tiled BAC clones for chromosome A9. Syntenic counterpart regions of the *A. thaliana* genome identified by sequence comparison were also represented in the colored boxes, along with position information. Red arrow heads indicate major gaps between scaffolds.

Color image of this figure appears in the color plate section at the end of the book.
applied to other *Brassica* genomic studies with the *B. rapa* genome as a reference and eventually contribute to accelerate the genomics progress, biological study, and molecular breeding in every *Brassica* crops.

**Acknowledgements**

We thank the many participants of the Korea *Brassica rapa* Genome Project. This work was supported by the National Academy of Agricultural Science (05-1-12-2-1, PJ006759, and PJ006691) and by the BioGreen 21 Program (20050301034438), Rural Development Administration, Korea.

**References**


O’Neill CM, Bancroft I (2000) Comparative physical mapping of segments of the genome of
Brassica oleracea var. alboglabra that are homeologous to sequenced regions of chromosomes 4 and 5 of Arabidopsis thaliana. Plant J 23: 233–243.


8

Exploring the Paradoxes of the Brassica Genome Architecture

Isobel A. P Parkin,1,* and Stephen J. Robinson1

ABSTRACT

The natural history of Brassica species is exemplified by the diverse morphological and biochemical phenotypes displayed among those cultivated as vegetables. Unique genome structures underlie this variation as the result of fluid genome dynamics consisting of repeated rounds of duplication, chromosome rearrangement events and subsequent diploidisation. However, this rapid Brassicae genome evolution has involved limited expansion of repetitive elements or transposition events characteristic of the evolutionary history of many plant tribes including the Triticeae. This has resulted in a remarkably high level of conservation among the Brassicaceae genomes. This chapter reviews the current knowledge of genome structure for Brassicaceae species at the chromosomal karyotype and at the physical sequence level. Such analyses have provided both tantalising insights into polyploid genome evolution and the identification of apparent contradictory forces that have played a role in shaping and maintaining these important plant genomes.

Keywords: genome evolution, polyploidy, transposon expansion, comparative mapping, genome architecture, paleobotany

8.1 Introduction

Grown worldwide vegetable Brassica are an important and highly diversified group of crops that are largely represented by Brassica oleracea and B. rapa species. In the western hemisphere, B. oleracea is the predominant species encompassing a number of valuable crops harvested for both their

1Agriculture and Agri-Food Canada, Saskatoon Research Centre, 107 Science Place, Saskatoon, SK, S7N 0X2, Canada.
*Corresponding author: isobel.parkin@agr.gc.ca
vegetative and floral structures including cabbage, broccoli, cauliflower, Brussels sprouts, etc. In Asia, *B. rapa* is the more commonly cultivated species owing to the great cultural and economic importance of Chinese cabbage. Generally, of less economic relevance are the *B. rapa* and *B. napus* cultivars that provide the root vegetable turnip and turnip greens and the vegetables mustards of *B. juncea* species.

Among the Brassicaceae there is a wide level of morphological variation found both between and within individual species including, inflorescence architecture, the size of roots and stems, patterning in leaf development and apical bud structures. This is most dramatically observed in the *B. oleracea* subspecies cauliflower (*botrytis*), broccoli (*italica*), cabbage (*capitata*), kale (*medullosa*), kohlrabi (*gongylodes*) and Brussels sprout (*gemmifera*) (Fig. 8-1). The natural history of the cultivated *Brassica* is interesting, as the genomes

![Figure 8-1](image-url)  
**Figure 8-1** Morphological variation among the Brassica vegetable species. A-G are *Brassica oleracea*, H is *B. rapa* and I is *B. napus*. Variation in head size and compactness is observed among Kale (A), Cabbage (B) and Brussels sprouts (C). Variation in inflorescence architecture is observed among Broccoli (D), Cauliflower (E) and Romanesco broccoli (F). Variation in stem and root size is observed in Kohlrabi (G), Bok Choy (H) and Swede/Rutabaga (I).
have undergone multiple rounds of polyploidisation and subsequent chromosomal rearrangement over a relatively short evolutionary period. The morphological variation so pronounced among *Brassica* species could result from changes in genome structure and gene regulation associated with the effects of polyploidisation. Studies focusing on these species offer the opportunity to investigate the impacts of polyploidy and the rapid evolution of phenotypic characteristics that has been observed (Lukens et al. 2004). Such research is facilitated by the close phylogenetic relationship of *Brassica* species to the plant model *Arabidopsis thaliana*, from which they are estimated to have diverged from a common ancestor approximately 15-24 Mya (Yang et al. 1999 Fig. 8-2). The extensive genetics and genomics resources currently available for *A. thaliana* can and have been exploited in the study of the *Brassica* species.

![Figure 8-2](image_url) Schematic detailing the significant events in the evolution of the modern day Brassica genomes, adapted from Rana et al (2004).

In this chapter what is known of the genomic structure of the vegetable *Brassica* species will be discussed, at both the macro-level of chromosome organisation and at the micro-level of the nucleotide sequence. Much of the work described has been advanced through the availability of the well annotated genome sequence of *A. thaliana*, which currently provides the optimal foundation for comparative genetics and genomics studies within the Brassicaceae (Arabidopsis Genome Initiative 2000). The landscape of Brassica research will change rapidly in the next few years with the imminent release of the first sequenced Brassica genome, that of Chinese cabbage (*B. rapa*) (Hong et al. 2008). The importance of the current knowledge of
Brassica genome organisation for interpreting and exploiting the promise of many sequenced Brassica genomes will be discussed.

8.2 Genome Dynamics of the Brassicaceae

8.2.1 Genome Content

In the 1930s the relationship between six *Brassica* crop species was proposed and later validated determining the eponymous U’s triangle (U 1935) (Fig. 8-3). The common vegetable *Brassicaceae* species are members of U’s triangle; *B. oleracea* (*2n* = 18, CC), *B. rapa* (*2n* = 20, AA) and *B. nigra* (*2n* = 16, BB) form the apices of the triangle and these species have fused in each pair wise combination to generate the larger and generally less diverse allopolyploid species *B. napus* (*2n* = 38, AACC), *B. juncea* (*2n* = 36, AABB) and *B. carinata* (*2n* = 34, BBCC). The Brassica genomes have undergone numerous rounds of chromosome doubling, nuclear hybridisations, and rearrangements, yet the genomes fall in the lower range of nuclear DNA content compared to other plant families (Johnston et al. 2005; Lysak et al. 2009). Among all reported Brassicaceae species none have been identified with a DNA C-value above 2.43 pg, in comparison to for example the cereal species of the Poaceae family, which range as high as 26.0 pg for a

![Figure 8-3](image-url)
The Triticale genome (Johnston et al. 2005; Lysak et al. 2009; http://data.kew.org/cvalues/). The low range of genome sizes, yet the acknowledged dynamic nature of genome evolution within the family offer an intriguing paradox. Some insight into the possible processes that maintain the low genome sizes has been gained from studies in Arabidopsis thaliana, which has one of the smallest known 1C DNA contents of 0.16 pg (Bennett et al. 2003) yet has evolved through paleopolyploidy (Blanc et al. 2003; Bowers et al. 2003; Henry et al. 2006). Although polyploid events and replication of repetitive sequences (see below) could lead to genome expansion, it appears that mechanisms to limit genome size are operating in the Brassicaceae family. For example, chromosomal rearrangements, perhaps as the result of non-reciprocal or unequal translocations, have been correlated with DNA loss in the evolution of Arabidopsis related species with reduced chromosome numbers (Yogeeswaran et al. 2005; Lysak et al. 2006). Interestingly the regulation of genome size does not appear to be influenced by selection (Lysak et al. 2009), thus the mechanisms and particularly the driving forces determining genome size remain to be completely elucidated.

8.2.2 Insights Offered by Comparative Genome Analysis

8.2.2.1 Comparisons Across the Brassica Species

Early cytological studies of chromosome pairing among Brassica species indicated the close relationship between the different species, particularly between the A genome of B. rapa and the C genome of B. oleracea (Attia and Röbbelen 1986a, 1986b). In resynthesised haploid B. napus lines where modern day B. rapa and B. oleracea genotypes were fused, homoeologous chromosome pairing was frequently recognised through the observation of multivalent chromosome formations. Yet, it was not until the application of molecular markers that it became apparent that such events were not the result of the forced intimacy, since homoeologous pairing between the A and C genomes occurred in natural B. napus lines at low but detectable levels (0.03–0.06 homoeologous recombination events per gamete) (Sharpe et al. 1995). Molecular markers revolutionised the study of the genome organisation in plants, and in particular the application of restriction fragment length polymorphism (RFLP) analysis indicated the dawn of comparative mapping studies. Such research has allowed the complex genome architecture of numerous plant species to be elucidated, visualising underlying duplication events and chromosome rearrangements, and establishing evolutionary relationships (Schmidt 2002). RFLP markers are particularly informative for such analyses, since each locus resolved represents a duplicate copy of the molecular probe used to identify the genetic loci, immediately establishing common evolutionary elements (see Chapter 4).
Numerous genetic linkage maps, including RFLP maps, have been developed for *Brassica* vegetable species: *B. oleracea* (Slocum et al. 1990; Kianian and Quiros 1992; Bohuon et al. 1996; Lan et al. 2000; Babula et al. 2003), *B. rapa* (Kim et al. 2006; Choi et al. 2007; Suwabe et al. 2008), *B. juncea* (Axelsson et al. 2000; Mahmood et al. 2003; Panjabi et al. 2008) and *B. napus* (Ferreira et al. 1994; Parkin et al. 1995; Sharpe et al. 1995; Lombard and Delourme 2001; Piquemal et al. 2005; Qiu et al. 2006). The limitation for comparative analysis among these species has been the restricted use of common marker sets; however, in the late 1990s comparisons between the three Brassica diploid genomes of U’s triangle became possible (Lagercrantz and Lydiate 1996; Parkin et al. 2003). These analyses were instrumental in developing our current understanding of chromosome architecture among the *Brassica* species, uncovering high levels of chromosome duplication and paradoxically considerable conservation of genomic segments across the genomes. The A and C genomes were found to be highly conserved with entire genetic linkage groups showing conservation of not only marker content but marker order (Fig. 8-4). The two genomes were separated by 16 detectable major chromosomal rearrangements, a number of which were associated with centromeric regions, suggesting that centromere fusion and fission are common forces leading to chromosome number deviations in *Brassica* species (Parkin et al. 2003). Comparisons of genetic linkage maps between all three *Brassica* diploid crop species indicate a larger number of rearrangements that separate the B genome from the A and C and none are shared with those that differentiate the A and C genomes (Lagercrantz and

---

**Figure 8-4** Cartoon representation of the relationship between the homoeologous regions of the *Brassica* A (grey) and C (black) genomes as defined in Parkin et al. (2003). The filled white circles show a subset of centromeres that have been positioned in the *B. oleracea* genome (A. Kelly and D. Lydiate, unpublished data).
Lydiate 1996; Panjabi et al. 2008). Strong conservation of extensive genome segments was also observed across all three genomes; despite this homology inherent difficulties have been found when attempting to transfer traits from the B genome into either of its related genomes (Struss et al. 1991; Chèvre et al. 1997) and homoeologous pairing was unexpectedly absent between the A and B genomes of resynthesised B. juncea (Axelsson et al. 2000). These data suggest that conservation of genome architecture may be restricted to the gross chromosomal structure and variation at the DNA sequence level is sufficient to prohibit pairing, this appears to be supported by the available expressed sequence tag (EST) data for B genome species (I. Parkin and A. Sharpe, unpubl. data).

The comparative genome analyses uncovered striking levels of genome duplication within all three Brassica genomes (Lagercrantz and Lydiate 1995; Parkin et al. 2003; Panjabi et al. 2008). These analyses noted that most maintained segments were arranged in three copies with varying levels of conservation, suggesting the Brassica species have evolved from a hexaploid ancestor likely formed from the fusion of three genomes of similar DNA content to the diminutive A. thaliana. The remnants of this ancestor have been further elucidated through comparisons with the plant model A. thaliana that have not only facilitated the resolution of evolutionary processes in this curious plant family, but also the exploitation of the substantial information and resources developed by the Arabidopsis community for Brassica crop improvement (Paterson et al. 2001).

8.2.2.2 Exploiting the A. thaliana Genome to Resolve Brassica Genome Architecture

Arabidopsis thaliana was adopted as a model for plant research largely due to its short generation time, small genome size and ease of transformation (Arabidopsis Genome Initiative 2000). Fortuitously for Brassica researchers although separated from the Brassicaceae tribe approximately 15–24 Mya (Yang et al. 1999; Bowers et al. 2003) A. thaliana has proved an excellent resource for facilitating Brassica research, conspicuously in the resolution of genome and gene structure.

Initial comparative analyses aligned targeted regions between the genetic linkage maps of Brassica species and A. thaliana (Osborn et al. 1997; Cavell et al. 1998; Parkin et al. 2002), identifying islands of gene content and gene order that were maintained to varying degrees across the evolutionary divide. The publication of the complete genome sequence for A. thaliana (Arabidopsis Genome Initiative 2000) marked a radical change in the resolution of comparative mapping within the Brassicaceae. The origin of genetic loci in Brassica species could be accurately positioned in the sequenced A. thaliana genome based on the strong sequence identity
observed between the species, 87% on average in genic regions (Lukens et al. 2003; Parkin et al. 2005). The most comprehensive analysis found that not only were the previously documented duplicated genomic segments preserved in the *A. thaliana* genome, but that the structure of the Brassica A and C genomes could be defined by approximately 21 of these regions (Parkin et al. 2005). These genomic segments ranged in physical length, from 2-9 Mb (the latter constituting almost 10% of the sequenced model genome) in the genome of *A. thaliana*, defining in some instances what appeared to be entire chromosome arms in the *Brassica* species, suggesting mechanisms are acting to maintain the structure of particular regions. Of note was the apparent conservation of chromosome landmarks such as centromeres, allowing the inference of as yet unmapped centromeric regions, both functional and archaic. These chromosome sites were also generally associated with breakpoints between conserved blocks and often correlated with numerous small segmental regions, which points to the fragile nature of such regions and their potential as rearrangement hotspots (Moore et al. 1997; Hall et al. 2006). Additional analyses have led to the resolution of at least 24 Brassicaceae blocks that can be replicated and rearranged to form the genomes of all members of the extended family (Schranz et al. 2006; Mandakova and Lysak 2008; Panjabi et al. 2008) (Fig. 8-5).

**Figure 8-5** The *Brassica oleracea* genome (C1-C9) generated from the conserved ancestral genomic blocks. The blocks are coloured according to the corresponding homologous region of *A. thaliana*: chromosome 1—light blue; chromosome 2—orange; chromosome 3—dark blue; chromosome 4—green; and chromosome 5—red. The names of the blocks are shown to the left of each Brassica linkage group according to Parkin et al. (2005) and their equivalent regions as described by Schranz et al. (2006) are shown to the right.

*Color image of this figure appears in the color plate section at the end of the book.*
The identification of the conserved blocks confirmed the hexaploid nature of the Brassicaceae ancestor, since invariably each of the blocks was found in three copies in each of the Brassica diploid genomes (Lysak et al. 2005; Parkin et al. 2005; Panjabi et al. 2008). This was also true for regions that had been replicated in the A. thaliana genome, itself an ancient polyploid (Blanc et al. 2003; Bowers et al. 2003; Henry et al. 2006), suggesting the duplications in A. thaliana pre-dated the triplication event that apparently differentiates the Brassicaceae tribe. However, some of the conserved blocks appear to have undergone further segmental duplication and potentially some have been lost subsequent to the triplication event.

Polyploidy is an important and prevalent phenomenon in plant evolution since an estimated 70% of flowering plants are suggested to have polyploidy in their lineages (Masterson 1994). The putative benefit of this evolutionary strategy lies in the duplication of gene content that facilitates the diversification of phenotype to provide selective advantage over the diploid relatives. The resolution of newly formed polyploids necessitates diploidisation of the nuclear genome, ensuring normal chromosome association, stable inheritance of genetic factors, restoration of fertility and ultimately the establishment of robust new species (Doyle et al. 2008). The mechanisms leading to diploidisation of these novel genomes are not completely resolved; however, two pathways are thought to predominate, massive chromosomal restructuring that suppresses pairing between non-homologous chromosomes or the rapid resumption of homologous pairing controlled by heritable genetic factors (Jenczewski and Alix 2004). It appears in the history of the Brassicaceae tribe that both mechanisms have played a role. In the Brassica diploids significant restructuring occurred after the triplication of the ancestral genome, making resolution of the ancestral karyotype problematic. However, there have been no observable chromosomal rearrangements since the recent generation of the Brassica allopolyploids (Parkin et al. 1995; Axelsson et al. 2000) and it has been proposed that they have inherited a genetic control of homologous pairing analogous to the Ph1 gene of wheat (Sears 1976).

Comparative mapping at the chromosomal level has provided insights into the level and extent of genome conservation and rearrangement within the Brassicaceae, these analyses have also suggested that each of the 24 genomic segments was not completely preserved in all three duplicated copies. It would be anticipated that the resultant level of gene redundancy might require deletion of genes copies or potentially allow gene divergence to stimulate sub- and neo-functionalisation of duplicates further contributing to the plasticity of the polyploid genomes (Lynch and Conery 2000; Lynch and Force 2000). Some of these questions are beginning to be answered with fine scale analyses, comparing the Brassicaceae genomes at the physical level and more recently at the nucleotide level.
8.3 Dissecting the Remnants of the Hexaploid Ancestor: Lessons from Studies of Microsynteny among Brassicaceae Genomes

In the absence of a complete Brassica genome sequence an improved knowledge of genome organisation has been drawn from targeted comparative analyses at the physical level between Brassica genomes and regions of *A. thaliana* containing genes of interest.

The study of genome architecture in polyploid species is complicated by the inherent duplication found, and in comparisons between *Brassica* species and their close relative *A. thaliana*, the underlying paleopolyploid nature of the model genome must be considered. Initial studies physically defined regions of the Brassica genomes by identifying overlapping sets of large insert clones or bacterial artificial clones (BACs) that were homologous to a target segment of the *A. thaliana* genome (O’Neill and Bancroft 2000; Rana et al. 2004; Park et al. 2005). Notably these first studies elucidated the structure of *B. oleracea*, *B. rapa* and *B. napus* segments homologous to a 222 Kb gene-rich region of *A. thaliana*, which was itself segmentally duplicated in the model. The design of these analyses provided further corroboration of the triplication event that defines the genomes of the Brassicaceae genomes and the fact that six regions were identified in both the A and C genomes compared to the duplicated region of *A. thaliana* placed the triplication event subsequent to the ancient duplications found in the *A. thaliana* lineage. It was apparent that the replication of the ancestral Brassica genome was followed by numerous changes in the microstructure with evidence of significant interspersed gene loss across the three genomic copies. Interestingly the microstructure of *B. napus* was found to closely mirror that of the diploid genomes indicating that maintenance of the contributing A and C genome within the newly formed allopolyploid nucleus was tenaciously controlled (Rana et al. 2004). The interspersed gene loss observed in comparison to *A. thaliana* is a common phenomenon in plant genome evolution and was originally speculated to be the result of unequal crossing over events (Bancroft 2001).

Sequencing of the six paralagous regions of the C genome identified in O’Neill and Bancroft (2000) has provided an illuminating glimpse into Brassica genome architecture and the processes driving change at the microstructure level in the *Brassicaceae* (Town et al. 2006). In *B. oleracea*, 66% of the genes that would be predicted based on triplication of the annotated *A. thaliana* gene identifiers were found in the collinear regions (Fig. 8-6). Invariably no evidence remained of the inferred missing genes, although in 10 instances gene fragments were identified in collinear regions, these data suggested that breaks in collinearity were largely the result of small interspersed deletion events, the mechanism for such a prevalent evolutionary device in the *Brassicaceae* remains to be elucidated. The presence
**Figure 8-6** Microsynteny between *A. thaliana* and three paralogous regions of the *B. oleracea* genome, adapted from Town et al. (2006). The figure shows interspersed gene deletion across the duplicate copies in *B. oleracea* and the compression/expansion of tandemly repeated genes. Arrows indicate the presence and orientation of a gene model. The horizontal lines indicate homology between gene models.
of 13 genes in *A. thaliana* not found in the sequenced Brassica genomic regions identified transposition of genes as an additional force driving changes to microstructure. Single gene transposition is becoming accepted as a significant phenomenon in plant lineage evolution and recent data suggests this has occurred at relatively high levels during the history of *A. thaliana* (Freeling et al. 2008). Large scale genome sequencing of plants has uncovered high levels of tandemly repeated genes, in particular 17% of *A. thaliana* genes were found arranged in tandem arrays (Arabidopsis Genome Initiative 2000). The study of disease resistance loci has suggested such arrays are dynamic in nature and evolve through homologous recombination and unequal crossing over (Michelmore and Meyers 1998). Town et al. (2006) identified not only the expansion and reduction of arrays in *B. oleracea* compared to *A. thaliana* (Fig. 8-6) but also determined that some arrays have become fixed over evolutionary time, perhaps as a function of their organisation, inverted versus parallel duplication, or potentially as a consequence of the selective advantage of the additive gene expression for those particular genes.

A similar analysis of paralogous regions of the *B. rapa* genome homologous to the *A. thaliana* genomic segments containing the *FLOWERING LOCUS C* (*FLC*) noted similar levels of collinear gene loss and transposition events (Yang et al. 2006) (see Chapter 7 for additional details). The authors also noted that some genes were modified through the insertion of small repetitive elements, which has been suggested as a mechanism for attaining novel protein functions (Bennetzen 2000). The significance of the changes in the microstructure of the Brassica genomes has yet to be aligned with gene function, although preliminary data studying gene expression levels across paralogous regions of *B. oleracea* have suggested that the multiple gene copies lead to differentiation of transcript abundance across tissue types (F. Iniguez-Luy, H. Wu, B. Underwood, M.L. Federico, J.S. Kim, J.C. Pires, R.M. Amasino, T.C. Osborn, C.D. Town, unpublished).

Mechanisms such as single gene transposition along with transposon-mediated genomic rearrangements such as duplications, excision/insertion events or retro-transposition are probably responsible for some of the genome specific differences observed between the A and C Brassica genomes. Data describing these differences is slowly being accumulated and it appears that the C genome has undergone genome expansion compared to the A genome that may partially explain the observed difference in genome size between the two *Brassica* species (Fig. 8-7). Figure 8-7 demonstrates the insertion of Brassica specific sequences in the C genome compared to both the Brassica A and the model *A. thaliana* genome (Bancroft et al. 2009). These sequences comprise transposable elements and transcribed regions.
reminiscent of the previously defined *B. oleracea* self-incompatibility locus (Fujimoto et al. 2006) suggesting transposition of a larger genomic segment or intra-chromosomal duplication.

![Figure 8-7 Comparative sequence analysis of a homologous region between *B. rapa*, *B. oleracea* and the model *A. thaliana*, based on data presented by Bancroft et al. (2009). This region shows expansion of the *B. oleracea* genome with Brassica specific sequences and the prevalence of transposition events in disrupting microsynteny across the genomes (circled gene models).](image)

The detailed analyses of paralagous Brassica regions has provided an estimate of gene content in the A and C genomes that is not dissimilar to that found in *A. thaliana*, specifically when comparing homologous regions: one gene per 4.8 Kbp in *B. oleracea* compared to 4.5 Kbp in *A. thaliana* (Town et al. 2006) and one gene per 3.7 Kbp and 3.4 Kbp in *B. rapa* and *A. thaliana*, respectively (Yang et al. 2006). The average gene length and structure are also remarkably conserved among the three species. Although the available data can only be considered an appetiser for the whole genome it is apparent that when the level of observed gene duplication is taken into account, the increased size of Brassica genomes compared to *A. thaliana* can not be explained alone by genome duplication events. The available global sequence data generated for both *B. rapa* (Yang et al. 2005) and *B. oleracea* (Zhang and Wessler 2004) provide some insights into the expansion process. Analyses of the two data sets indicate that while both genomes contain a similar level of protein coding genes, *B. oleracea* contains a higher level of characterised transposable elements (20%) than its close relative *B. rapa* (12%). The role of transposable elements in genome evolution is a rapidly developing area of research particularly with the current exponential increase in available genome sequence data for plant species.
8.4 Contribution of Low Complexity Sequences to the Brassiceae Genomes

In the plant kingdom, repetitive elements are thought to be largely responsible for the extreme variation in genome size observed among angiosperm species, this is exemplified by the rapid increase in the nuclear content of maize through the accumulation of retrotransposons that contribute 50–60% of the genome (Bennetzen et al. 1998). In most eukaryotic organisms repetitive sequences are not evenly distributed across the genome but are found to cluster in heterochromatic regions, with telomeres and centromeres containing large numbers of diverse repeat sequences that contribute to the maintenance of chromosome stability and normal segregation (Henikoff et al. 2001). In Brassica species due to difficulties in resolving the karyotype, it is only recently that the positions of the centromeric regions have been accurately identified and only in the case of B. napus has this information been linked to the genetic maps, which has confirmed the conservation of centromere positions across Brassiceae species (Lim et al. 2007; Pouilly et al. 2008).

8.4.1 Pericentromeric and Centromeric Repeats

The centromere is an essential structure that ensures the faithful segregation of chromosomes at cell division, and consists largely of highly repetitive DNA sequences, specifically tandem satellite repeats (Jiang et al. 2003). The role of these tandem repeats has yet to be fully elucidated but they are believed to recruit DNA binding proteins during chromosome condensation facilitating the replication and segregation of the chromosomes (Csink and Henikoff 1998). Centromeric repeats have been characterized in a number of plant genomes and are notable for their extreme diversity with little sequence homology found across plant families, indicating they are rapidly evolving, which contrasts starkly with their fundamental role. The centromeres of B. rapa and B. oleracea are composed of 176-bp satellite repeat sequences (CentBr), which are not homologous with the centromeric repeats of either the model A. thaliana or the more closely related species, B. nigra (Harrison and Heslop-Harrison 1995; Lim et al. 2005; Lim et al. 2007). Further, the CentBr repeats in the B. rapa genome have been shown to represent two classes which are chromosome-specific, one being found on eight and the second on only two chromosomes (Lim et al. 2007) (see Chapter 7 for additional details). Sequencing of peri-centromeric regions is resolving the organisation of these complex structures and has identified a number of transposable elements among the satellite repeats that have been conserved across the Brassiceae species (Lim et al. 2007).
8.4.2 Transposable Elements

Transposable elements have undergone a renaissance from selfish junk elements to active facilitators of plant genome evolution; through gene capture, exon shuffling, gene transposition and the generation of novel polymorphism (Bennetzen 2005; Morgante 2006). A diverse range of transposable elements have been described in plants and are distinguished based on their mode of replication, either facilitating movement via an RNA intermediate (Class I) or via a direct DNA excision-insertion mechanism (Class II). Genome sequence analysis has identified that 20% and 12% of the *B. oleracea* and *B. rapa* genomes are contributed by transposable elements, respectively (Zhang and Wessler 2004; Yang et al. 2005).

The majority of transposable elements found in the vegetable Brassica genomes were classified as Class I retrotransposons, comprising 70-89% of the total identified elements. The accurate classification of transposable elements has only been completed for *B. oleracea*, which allowed comparisons with *A. thaliana* elements and provided insights into the evolution of these DNA molecules across the Brassicaceae genomes (Zhang and Wessler 2004). The *B. rapa* genome sequencing project is providing insights into the prevalence of these elements in the A genome and the analyses thus far is described in Chapter 7. The class I transposable elements are subdivided in two, those possessing long terminal repeats (LTRs) and those without LTRs that terminate with a polyadenaline tract, which are further subdivided into long interspersed nuclear elements (LINEs) or short interspersed nuclear elements (SINEs).

In *B. oleracea*, the LTR-elements predominate the Class I type and are composed of families of *copia*-like and *gypsy*-like retrotransposons, in each case where related lineages were found in both the model and crop genome they were more prevalent in *B. oleracea*. All the *copia*-like elements found in *A. thaliana* were represented in *B. oleracea*; however, one family of elements was specific to the Brassica C genome. The *gypsy*-like elements appeared more evolutionary disparate with less than half the identified families being common to the two species and the remainder having emerged after their divergence from a common ancestor. LINEs are recognized as the most ancient class of retrotransposons in plants which was reflected in the high level of conservation observed between *B. oleracea* and *A. thaliana*. This was in stark comparison to SINEs where a novel class of elements was identified in *B. oleracea*, BoS, that was distinguished as the most abundant class of SINEs in the Brassicaceae and was found to be composed of an incredibly diverse range of related elements a proportion of which had apparently evolved through chimerism of more ancient elements (Zhang and Wessler 2005).
Class II transposons were much less abundant than their class I counterparts, approximately 6% of the *B. oleracea* genomic content was comprised of class II elements. Interestingly, two families of CACTA-like elements in *B. oleracea* appear to have undergone exponential expansion (*BoC1* and *BoC2*), contributing one third of the total class II elements. The expansion of class II elements in plant genomes is a relatively new observation since their preference for insertion in gene-rich regions would suggest increased density would be limited. It has yet to be determined if the expansion of class II elements in *B. oleracea* is specific to the C genome; however preliminary data have identified a smaller proportion of DNA-type transposable elements in available *B. rapa* sequence data (Hong et al. 2008).

The compact nature of the *A. thaliana* genome is reflected in the low abundance of transposable elements (~ 10% of the total genome), which is in contrast to that found in the Brassica lineages. The comparative analyses of such elements between *B. oleracea* and *A. thaliana* established the age of each element lineage, the level of replication in the two related genomes and determined their potential mode of evolution (Zhang and Wessler 2004). The difference between the genomes could result from a concerted loss or the prevention of expansion of transposable elements by an unknown control mechanism in the *A. thaliana* lineage. The latter hypothesis appears more likely since nearly all subfamilies of transposons have been found in both the Brassica and *A. thaliana* lineages and there is no single transposon family that has undergone significant expansion in the *A. thaliana* genome. In comparison to other plant families, all the Brassicaceae species studied have maintained remarkably low levels of transposable elements. In maize four families of LTR elements alone contribute 32% of the genomic content, in contrast to *B. oleracea* although there was a multiplication of both class I and class II elements there has been no amplification of specific retrotransposon families observed, instead many lower abundance families are maintained. Due to the paleoploid nature of the Brassica genomes this result would not be anticipated, since the genomic redundancy should provide an environment for transposons to flourish. These data suggest that mechanisms are regulating the transposition potential within the Brassicaceae genomes and as expression data accumulates it will be possible to identify active transposons from those no longer able to replicate, which may shed light on the genetic processes at play.

### 8.5 Conclusions

A plant’s genome architecture partly determines the plasticity of a species and in affect controls its ability to respond to diverse environmental pressures. Polyploid species appear to have a selective advantage since
their inherent genetic redundancy allows for rapid diversification of gene content and gene expression. Indeed the Brassica vegetable crops have evolved from a common hexaploid Brassicaceae progenitor to provide a diverse array of species, which vary in morphological, biochemical and geographical profiles. The release of the genome sequence for B. rapa is anticipated for 2010 and should begin to answer some of the remaining questions regarding Brassica genome structure. This information could finally provide a comprehensive picture of the level and distribution of genome duplication and potentially facilitate insights into the evolutionary mechanisms that have driven the generation of this valuable and diverse group of species. The identification of the estimated 50,000 genes in each of the Brassica diploid genomes will allow the role of duplicate gene copies in defining the overwhelming phenotypic diversity to be fully described. The escalating ability to capture sequence data accurately and economically should lead to the genomes of a number of Brassicaceae and related species being released in the next few years, predicting an exciting and revelatory period for Brassica and polyploid genome research.

References


Exploring the Paradoxes of the Brassica Genome Architecture


Exploring the Paradoxes of the Brassica Genome Architecture


The Decoding of Gene Functions—Transcriptomics Tools Toward Molecular Physiology and Breeding

Agnieszka Ludwików,1,* Krzysztof Krasowski,1 and Leo Zeef2

ABSTRACT

Brassica species are particularly important vegetable crops, that have been studied in many aspects of plant biology including seed growth, development, disease resistance and abiotic stress tolerance. In the past decade new genomic resources have been developed and subsequently used for a generation of tools to enable genome–wide screens in Brassica. Transcriptomic platforms have been established to facilitate the identification of genes underlying complex traits. Despite the substantial progress that has been made toward uncovering the network of gene functions in conditions investigated, much work still needs to be done to completely characterize associations between gene expression patterns and the genotypes in breeding populations. This chapter presents the current status of transcriptomic resources and efforts in determining the functions of genes in Brassica species.

Keywords: Brassica vegetables, genomics resources, microarrays, comparative genomics, stress tolerance

1Department of Biotechnology, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland.
2School of Biological Sciences, University of Manchester, Manchester M139PT, UK.
*Corresponding author: ludwika@amu.edu.pl
9.1 Introduction

For the last decade plant functional genomics was developed based on model organisms such as *Arabidopsis thaliana* and more recently *Oryza sativa*. The knowledge gathered is now extensively used to study other plants including *Brassica* crop species. *Brassica* species have been studied in many aspects: seed growth, development, disease resistance and abiotic stress tolerance (van Poecke and Dicke 2002; Tuberosa and Salvi 2006; Broekgaarden et al. 2007, 2008). Vegetable Brassicas are recognized as a valuable source of vitamins, nutrients and anticancer compounds (Paterson et al. 2001). The large effort to bridge the gap between model plants and crop species functional genomics studies leveraged the development of many genomics resources for *Brassica*. Several tools were established to decipher gene functions (van Poecke and Dicke 2002; Tuberosa and Salvi 2006; Broekgaarden et al. 2007, 2008). These include expressed sequence tag (EST) collections, cDNA libraries, comprehensive mutant collections and DNA microarrays (Lim et al. 1996; Kwak et al. 1997). Special attention has been given to the generation of transcriptomic platforms, which are now released to the community and well supported. So far the *Brassica* transcriptome has been analyzed to address both agricultural and fundamental biological questions (Paterson et al. 2001; Kwun et al. 2004; Li et al. 2005; Soeda et al. 2005; Joosen et al. 2007; Kang et al. 2008; Lee et al. 2008; Xiang et al. 2008; Sarosh et al. 2009). The available microarray experiments uncovered genes with diverse functions that may act in regulation of embryogenesis and seed development (Lim et al. 1996; Soeda et al. 2005; Li et al. 2005; Kang et al. 2008; Xiang et al. 2008), stress response, and stress tolerance (Hammond et al. 2005; Vanderauwera et al. 2007; Zhou et al. 2007; Broekgaarden et al. 2008; Lee et al. 2008). Various analytical approaches were employed to facilitate the identification of new *Brassica* specific genes that respond to biotic and abiotic stimuli (Hammond et al. 2005; Broekgaarden et al. 2008; Lee et al. 2008). The available microarray studies revealed a large variety of functional assignments based on the conserved expression patterns of othologous genes. The microarray approach also has the potential to elucidate regulatory pathways controlling important traits, thereby increasing the effectiveness of plant breeding. This chapter presents the current status of transcriptomic resources and efforts in determining the functions of genes in *Brassica* species.

9.2 Functional Genomics Resources for *Brassica* species

9.2.1 Transcript Profiling Tools

To determine the relationships between gene structure and gene expression several techniques and platforms have been developed. These include
methods such as cDNA library and EST production, serial analysis of gene expression (SAGE), massively parallel signature sequencing (MPSS), differential display PCR, suppression subtractive hybridization (SSH), real-time quantitative PCR, and microarray-based techniques. Broadly, transcript profiling techniques can be classified into two categories: sequencing-based and hybridization-based methods (Liu et al. 2007). With the exception of cDNA libraries, other techniques even if they are considered routine, are not frequently utilized for the estimation of transcript abundance in vegetable Brassica species. For example MPSS that leads to the faster discovery of novel single nucleotide polymorphisms (SNPs) and novel transcripts have not yet been used in vegetable Brassica research. The SAGE technique has been successfully applied to identify genes that are differentially expressed during seed development. However, among Brassica species this technique was applied just in the case of B. napus (GEO records: GSE14313; GSE7204; GPL8033; GPL4919). Suppression subtractive hybridization (SSH), one of the most popular methods for generating subtracted cDNA or genomic DNA libraries, was used by Cramer et al. (2006) to identify genes involved in Alternaria brassicicola—Brassica oleracea interaction.

The most popular approach to gene discovery in Brassica studies was the generation and screening of EST and cDNA libraries. This strategy provides global views of gene expression levels based on EST frequency. Currently, cDNA libraries have been used for identifying the genes associated with a physiological process or tissue specific expression (Lim et al. 1996; Kwak et al. 1997; Lim et al. 2000). For the functional genomics of Brassica campestris, a guard cell cDNA library has been constructed to gain an insight into the functions and development of this cell type (Kwak et al. 1997). The identified ESTs showed similarity to genes of different plant species including the Brassicaceae as well as non-plant genes. Several ESTs were also related to guard-cell functions such as regulation of stomatal movements. In other studies cDNA library-based transcript profiling was used with the aim to uncover the mechanism of the heading process in B. rapa. This study showed that majority of the identified ESTs share homology to the putative protein sequences of known biological functions and the highest score obtained for A. thaliana sequences (Rui-Juan et al. 2004). Furthermore, the partial sequencing of 1,216 randomly selected cDNA clones from B. campestris flower buds resulted in the identification of 904 cDNA clones (Lim et al. 1996). Nearly 50% of them had sequence homology to functionally defined genes. Using a similar methodology Lim et al. (2000) generated ESTs from different organs at various developmental stages of Brassica rapa L. ssp. pekinensis. This study can be applicable for the comprehension of the developmental processes of the Brassica plants and for the comparative analysis of Arabidopsis genome structure.
Efforts to generate ESTs are now being extended. Additional organ- or developmental stage-based cDNA libraries *B. oleracea* and *B. rapa* have been recently constructed. Continuously acquired data is expanding our knowledge about genome size and gene function in *Brassica*. Existing *B. napus* libraries cover a large amount of information regarding organ specific expression and development. Available cDNA libraries and ESTs were further used to facilitate the construction of transcriptomic microarrays (Brassica Genome Gateway). The microarray-based approach for gene expression profiling has shown enormous potential and is now the most widespread method (technique) applied in transcriptomes research. Microarray studies have been used in several *Brassica* species to better understand the changes in gene expression during growth and development, pathogenesis, abiotic stresses and many biological processes. These days, microarray-based gene expression monitoring is possible through high-density oligo or full-length cDNA *Brassica* arrays, *Arabidopsis* ATH1 GeneChip arrays and less frequently, custom low-density macroarrays utilizing *Arabidopsis* cDNA/EST as a probe. Unfortunately, the most recently developed BeadArrays (Kuhn et al. 2004) and high-throughput screening qPCR (e.g. TaqMan Gene Expression Assays) were not developed for *Brassica* research.

### 9.2.2 Microarray Resources

The development of a range of expression profiling tools for *Brassica* species have been achieved in the frame of different international networks. Five technology platforms have been developed including the Agilent eArray, the Combimatrix array, the NimbleGen array, and custom cDNA arrays spotted on nylon membrane (macroarray) or glass support. As oligo design supports sequence specificity, three out of four types of available *Brassica* microarrays have been designed up to 60-mer oligo probes (Table 9-1). The Agilent eArray, the Combimatrix and the NimbleGen arrays were designed based on available EST deposited in GenBank and large *Brassica* databases.

<table>
<thead>
<tr>
<th>Technology platform</th>
<th>Probes</th>
<th>Probe length</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agilent 2x 105K</td>
<td>91,854</td>
<td>60-mer</td>
<td>for use with all <em>Brassica</em> Species</td>
</tr>
<tr>
<td>Combimatrix 94K</td>
<td>~ 90,000</td>
<td>35-mer</td>
<td>for use with all <em>Brassica</em> Species</td>
</tr>
<tr>
<td>NimbleGen KGBP 24K</td>
<td>331,384</td>
<td>60-mer</td>
<td><em>B. rapa</em> microarray</td>
</tr>
<tr>
<td>NimbleGen <em>B. rapa</em> 300K</td>
<td>360,312</td>
<td>60-mer</td>
<td><em>B. rapa</em> microarray</td>
</tr>
<tr>
<td>Seed cDNA array</td>
<td>10,642</td>
<td>cDNA</td>
<td>for use with all <em>Brassica</em> Species</td>
</tr>
<tr>
<td>Custom cDNA arrays</td>
<td>cDNA</td>
<td></td>
<td>for use with all <em>Brassica</em> Species</td>
</tr>
<tr>
<td>Affymetrix ATH1</td>
<td>25-mer</td>
<td></td>
<td><em>Arabidopsis</em> GeneChips arrays</td>
</tr>
</tbody>
</table>
The Decoding of Gene Functions

95K microarray generated by the J. Craig Venter Institute (JCVI), The John Innes Centre (JIC) and Cogenics is a validated, one probe per gene, 91,854 60-mer oligonucleotide array synthesized in situ using Agilent Technologies’ SurePrintTM technology in 2 × 105K format (an inkjet printing method). The updated description of the array such as EST assemblies, gene ontology distribution of ESTs associated with the array as well as estimations of species-specific genes are available on the Brassica Gateway website. Besides Brassica Unigenes, the array includes both negative controls (plant and human genes) and positive controls (insect housekeeping genes). The CombiMatrix 94K microarray system is the only one that offers ~90,000 Brassica probes on up to three times re-usable slides (Ma et al. 2004; Xiang et al. 2008).

Two arrays have been designed at the NimbleGen for functional genomics of Brassica rapa: 300K Microarray v2 and KGBP 24K Microarray with 47,548 and 24,963 unigenes, respectively. The 300K Microarray consists of 331,384 of 60-mer probes. Probes were synthesized in situ using a Digital Micromirror Device (DMD). An advantage of the NimbleGen platform is the relatively low cost of microarray preparation due to the absence of pre-existing masks as it occurs in Affymetrix GeneChips. Seven probes that cover 150 bp in the 3’ region of the gene represent each gene. The array includes genes from chloroplast, mitochondria and selection marker genes. Alternatively, KGBP 24K Microarray consists of 360,312 probes duplicated in 2 blocks. Six probes represent each gene and ~20% of genes with an unknown direction have 12 probes designed as follows: 6 probes in sense and 6 in an antisense direction. Sequences of probes and unigenes are available at BrGD (http://www.brassica-rapa.org). Other resources are also available. For example, SIPPE Brassica napus 8.1K cDNA Microarray (v1.0), NEFU Brassica rapa 3k v1.0A as well as 10,642 Brassica Seed cDNA array were preliminary used for the identification of the genes involved in seed development processes or unique traits (Xiang et al. 2008). Additionally, for a comparative analysis ATH1 Genechip or other custom Arabidopsis arrays can be used. Originally, the A. thaliana ATH1-121501 Genechip array was used to study gene expression in Brassica oleracea (Hammond et al. 2005). Moreover, a 60k Brassica Affymetrix array, based on the JCVI assemblies, should be also available soon (www.brassica.info).

9.2.3 Microarray Analysis and Tools

One can be easily confused by the wide variety of microarray analysis tools currently available. However, if one looks beyond the specific software packages and algorithms, commonalities are apparent in the steps that need to be taken during microarray analysis (Allison et al. 2006). In this section we will present steps that should be performed and give some examples of
software tools available. Since *Brassica* studies are likely to be performed on a variety of microarray platforms (Hammond et al. 2005; Katari et al. 2005; Soeda et al. 2005; Esmon et al. 2006; Hudson et al. 2007; Xiang et al. 2008) we have kept the information applicable to most microarray platforms.

### 9.2.3.1 Experimental Design

Knowing the possible sources of error is important. Since microarrays do not report absolute quantification but only comparative hybridization, this places great importance on the control to which the treated sample must be compared. Potential sources of difference between the treated sample and control, extraneous factors such as age differences of the subjects (when that is not the factor under investigation), or the manner and time of harvesting, processing RNA samples or running arrays, should be eliminated or at least brought to a minimum. To have confidence in differential expression of genes being measured by microarray, replication should be performed and this should be biological replication rather than technical, where the same RNA is placed on two or more arrays. Pooling is an acceptable way of improving statistical power without hugely increasing costs, however, replicates should still be performed. Pooling destroys information about the variability of RNA expression between individuals so verification of results should be done on unpooled samples (using quantitative RT-PCR for example).

### 9.2.3.2 Quality Control of Microarrays

Following the need to control extraneous factors that were emphasized in the previous paragraph, one seeks consistency between the arrays when it comes to quality analysis. There may be particular target metrics that the microarray manufacturer recommends, but in an experiment it is important to check that metrics like median intensity, distribution of intensities and background intensities are consistent. Small differences can be compensated for by normalization methods as we shall see in the next paragraph, but there is a limit to how effective these corrections can be. Checks should be done for technical artefacts and special defects on the array such as blobs and spot irregularities.

### 9.2.3.3 Image Processing and Normalization

For a given array type several methods may be available for estimation of the amount of RNA from fluorescent array images, while trying to minimize the extraneous variation that occurs owing to technical artefacts. The Bioconductor package, based on the R programming language, is an
excellent open source and open development software project making these methods freely available (Gentleman et al. 2004). For example, when analyzing Nimblegen arrays, the “oligo” library can be used to perform the RMA method of image processing and normalization that is commonly used for analysis of Affymetrix arrays (Bolstad et al. 2003). Image processing and normalization will produce a spreadsheet of expression values for each gene. The data are commonly transformed into logarithmic scale (typically log base 2) for further analysis.

9.2.3.4 Quality Assessment of the Experiment

Performing a principal components analysis (PCA) is a valuable way of assessing the performance of the experiment. If replicates have been performed then grouping of replicates indicates that reliable differential expression has occurred between the conditions studied. If on the other hand a random distribution of the samples in PCA space is found then great caution needs to be exercised in the statistical analysis described in the next paragraph (when thousands of genes are analyzed at once there will always be some genes that appear differentially expressed in a statistically significant way due to random chance events). When more than one factor is being studied, for example drought and aphid resistance, a PCA can indicate the relative strengths of these effects on gene expression.

9.2.3.5 Inference

A statistical test is performed on a gene by gene basis to assess if differential expression has occurred. Using fold change alone is not acceptable. A parametric test such as the Student’s t-test or ANOVA can be used on the data in logarithmic scale. Prior to this, removal of defective spots may be required, depending on the microarray platform used. Because of costs a typical microarray experiment contains a smaller number of replicates than is optimal. Variance shrinkage methods are useful in making use of the parallel-testing nature of microarrays to bolster statistics (Cui et al. 2005). The parallel-testing nature of microarrays leads to a different statistical problem when it comes to choosing a set of genes that show differential expression (referred to as a genelist). Selecting a genelist “with p values less than 0.05” leads to a multiple-testing error and so a false-discovery correction should be applied (Storey 2003).

9.2.3.6 Genelist Analysis

Once a set of statistically significant differentially expressed genes has been identified, the next challenge is to organize the list of genes. Analyzing the genes into classes, often based on Gene Ontology categories, is helpful in
this regard. An intermediate step of comparative genomics may be required to make use of the information available in the Arabidopsis model species.

9.2.4 Genomic Projects and Databases

Numerous laboratories collaborate in several consortia to build resources that facilitate gene function discovery in Brassica crops. These resources include: a BBSRC Innovations in Crop Science Project headed by the AdVaB consortium, the Multinational Brassica Genome Project (MBGP), the Multinational Brassica rapa Genome Sequencing Project (BrGSP), and the Defra funded Oilseed Rape Genetic Improvement Network (OREGIN) (Brassica.info). In total, there are 10 Brassica Genome Projects: four for B. napus (EST project at Genoplante; Plasmid genome at the National Agricultural Research Center for the Hokkaido Region, Japan; Genome sequencing project at John Innes Centre and Beijing Genomics Institute), two for B. rapa (Genome sequencing at BrGSP and MBGP), two for B. oleracea (Genome sequencing at TIGR; physical map and comparative genomics at BBSRC) and single for B. juncea and B. nigra (both are part of MBGP). These organizations through their independent and collective activities develop Brassica resources, integrate technologies and build technical support for their practice usage. Ongoing sequencing projects and other activities force the Brassica community to develop comprehensive databases and public repositories (Table 9-2). The biggest public repository database at NCBI, in cooperation with other Brassica-related resources, including ASTRA, BAGI, BGG, CSCL, and MBGB, have accumulated information for the Brassica species including 651 libraries, 19,343 nucleotide sequences, 850,713 ESTs, 970,602 GSSs, 2,066 probes, 10 genome projects, 46,855 UniGene and no SNP and dbGaP (Genotype and Phenotype data) records and many others. For vegetable Brassica species GeneBank deposits information about 144 cDNA libraries, 11,374 nucleotide sequences, and 249,090 ESTs records (Table 9-3).

As indicated by GeneBank, the sequencing of 118 B. rapa and 26 B. oleracea cDNA libraries generated 246,224 expressed sequence tags. Additional 2,866 ESTs are available for B. carinata, B. juncea and B. nigra. Several specific B. rapa cDNA libraries were also constructed, including whole plants library treated by light-chill or NaCl stress, calli in organogenesis stages on the tissue culture media library as well as cDNA library from leaves infected by the pathogen Pectobacterium carotovora subsp. carotovora (NCBI, BrGP). These libraries provide very useful information in relation to the biology of the stress response and regeneration of plants. Furthermore, ESTs from the B. rapa and B. oleracea libraries can be grouped into 20,114 UniGene
Table 9-2 Genomics resources for Brassica species.

<table>
<thead>
<tr>
<th>Databases and public domains</th>
<th>Website URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brassica Genome Gateway</td>
<td><a href="http://brassica.bbsrc.ac.uk/">http://brassica.bbsrc.ac.uk/</a></td>
</tr>
<tr>
<td>Brassica Genomics</td>
<td><a href="http://www.brassicagenomics.ca/">http://www.brassicagenomics.ca/</a></td>
</tr>
<tr>
<td>Brassica.info</td>
<td><a href="http://www.brassica.info/">http://www.brassica.info/</a></td>
</tr>
<tr>
<td>OREGIN</td>
<td><a href="http://www.oregin.info/">http://www.oregin.info/</a></td>
</tr>
<tr>
<td>WHRI Brassica Genetic Resources</td>
<td><a href="http://www2.warwick.ac.uk/fac/sci/w">http://www2.warwick.ac.uk/fac/sci/w</a> hilar/research/gre/b brassica/</td>
</tr>
<tr>
<td>The John Innes Genome Laboratory</td>
<td><a href="http://www.jicgenomelab.co.uk/libraries/b">http://www.jicgenomelab.co.uk/libraries/b</a> brassica.html</td>
</tr>
<tr>
<td>Brassica oleracea Genome Project</td>
<td><a href="http://www.tigr.org/tdb/e2k1/bog1/">http://www.tigr.org/tdb/e2k1/bog1/</a></td>
</tr>
<tr>
<td>Brassica European Database</td>
<td><a href="http://documents.plant.wur.nl/cgn/pgr/brasedb/default.htm">http://documents.plant.wur.nl/cgn/pgr/brasedb/default.htm</a></td>
</tr>
<tr>
<td>Brassica Working Group</td>
<td><a href="http://www.ecpgr.cgiar.org/Workgroups/brassica.htm">http://www.ecpgr.cgiar.org/Workgroups/brassica.htm</a></td>
</tr>
<tr>
<td>AAFC Brassica / Arabidopsis Genomics Initiative (BAGI)</td>
<td><a href="http://brassica.agr.gc.ca/index_e.shtml">http://brassica.agr.gc.ca/index_e.shtml</a></td>
</tr>
<tr>
<td>Brassica rapa Genome Project</td>
<td><a href="http://www.brassicam-ra">http://www.brassicam-ra</a> pa.org/BGP/NC_brgp.jsp</td>
</tr>
<tr>
<td>Brassica rapa EST and Microarray database (BrEMD)</td>
<td><a href="http://www.brassicam-ra">http://www.brassicam-ra</a> pa.org/BrEMD/</td>
</tr>
</tbody>
</table>

resources (putative genes; Table 9-3). Having these estimations, we can speculate that the NCBI UniGene records cover less than 30% of the genes in the B. rapa genome (50,000 estimated transcripts) and 7% of B. oleracea genome (80,000 estimated transcripts), respectively. It is important to note that, the Brassica rapa Genome Project (BrGP) headed by two teams Korea Brassica Genomics (KBGP) and the Biogreen21 Program (BG21) reported a detailed analysis of 127,144 ESTs from the 26 different cDNA libraries of Brassica rapa. Following this information; the investigated EST sets generates 32,395 Unigene sequences that cover around 65% of the genes in B. rapa genome. The most current list of Brassica databases and related resources can be found on Brassica Genome Gateway webpage (http://brassica.bbsrc.ac.uk/) and Brassica.info (http://www.brassica.info/).

Table 9-3 Brassica resources.

<table>
<thead>
<tr>
<th>Brassica species</th>
<th>ESTs</th>
<th>cDNA libraries</th>
<th>Nucleotide sequences</th>
<th>Unigene records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brassica napus</td>
<td>656,145</td>
<td>542</td>
<td>11,871</td>
<td>27,143</td>
</tr>
<tr>
<td>Brassica rapa</td>
<td>194,741</td>
<td>140</td>
<td>4,242</td>
<td>14,606</td>
</tr>
<tr>
<td>Brassica oleracea</td>
<td>63,201</td>
<td>26</td>
<td>5,478</td>
<td>5,617</td>
</tr>
<tr>
<td>Brassica carinata</td>
<td>2,500</td>
<td>-</td>
<td>101</td>
<td>-</td>
</tr>
<tr>
<td>Brassica juncea</td>
<td>5,393</td>
<td>-</td>
<td>1,034</td>
<td>-</td>
</tr>
<tr>
<td>Brassica nigra</td>
<td>1,814</td>
<td>-</td>
<td>624</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>923,794</td>
<td>708</td>
<td>23,350</td>
<td>47,366</td>
</tr>
</tbody>
</table>

Resources provided by the NCBI Entrez database; update for April 2010.
There was a significant acceleration of *Brassica* functional genomic research over the past few years. Unfortunately, it was not followed by the integration of public gene expression databases for *Brassica* species. The development of microarray platforms allowing specific in-species *Brassica* analyses as well as cross-species comparisons would advance the *Brassica* gene expression studies. Currently, the *Brassica rapa* EST and microarray database (BrEMD) store 10 microarray datasets and allows simple microarray data analysis as well as keyword search. There are only 18 different microarray experiments deposited in the existing repositories including the Gene Expression Omnibus database (GEO), EBI-ArrayExpress, ArrayExpress and NASCarrays. Some datasets are deposited in two databases. Fifteen microarray experiments are in GEO database: 10 experiments for *B. napus*, three for *B. rapa* and four for *B. oleracea*. ArrayExpress includes six microarray experiments: two for each: *B. napus*, *B. rapa* and *B. oleracea*. Additionally, NASCarrays has deposited three microarray experiments for *B. oleracea*. Beyond data analysis results this databases store information about data structure, protocols, and annotations in the MIAME compliant format.

In contrast to microarray data, comprehensive EST databases and online annotation tools have been well developed. The *Brassica* Expressed Sequence Tag Portal at Brassica Genomics ([http://www.brassicagenomics.ca/](http://www.brassicagenomics.ca/)) contains ESTs from four *Brassica* species (*B. napus*, *B. rapa*, *B. oleracea* and *B. carinata*), provides ESTs annotation and information about EST sequence homology. EST analysis tool at *Brassica* Genome Gateway ([http://brassica.bbsrc.ac.uk/](http://brassica.bbsrc.ac.uk/)) allows both BLAST and EST ID search for Brassica 95k microarray v1 array. In the output it shows the homology search best hit (gene name, protein ID) and EST assembly components. Alternatively, *Arabidopsis* BLAST search at TAIR ([www.arabidopsis.org](http://www.arabidopsis.org)) also provides an additional, advanced and well organized tool for gene annotation and metabolic pathway analyses in *Brassica*.

### 9.3 Comparative Genomics

The genomes of *Brassicaceae* are large and complex (Paterson et al. 2001; Paterson 2005; Xiang et al. 2008). Comparative analyses of chromosome segments have shown the conservation of gene order in *Arabidopsis* and *Brassica* (Paterson 2005; Ziolkowski et al. 2006). Given the high genomic collinearity, the comparative genomics approach gives an opportunity to transfer information from model species to crops. Comparative genomics improve our understanding in areas of phylogenomics (genomic data is used to construct more robust phylogenies), structural genomics (genome rearrangements: segmental duplications, deletions, insertions, transpositions, polyploidy), annotation
and selection/evolution of traits (Paterson 2005; Chen and Ni 2006; Schranz et al. 2006; Ziolkowski et al. 2006; Schranz et al. 2007). In the past few years a comparison of *A. thaliana* relatives namely *B. oleracea* and *B. rapa* helped develop our knowledge about Brassicaceae.

### 9.3.1 *Arabidopsis thaliana*—*Brassica oleracea*

Comparative genomics is an effective method to annotate the sequenced genome by identifying genes, regulatory elements and improving gene structure (Colinas et al. 2002; Hammond et al. 2005; Katari et al. 2005; Esmon et al. 2006). A comparative approach has a great capacity to complement current annotation. An excellent example of the comparative genomics strategy is the transcriptome analysis of *Brassica oleracea*, by using an ATH1 Genechip array designed for *A. thaliana* (Hammond et al. 2005). Importantly, to improve the reliability of *Brassica* hybridization results the authors have identified probe-pairs that hybridized to the *B. oleracea* genomic DNA on the basis of the perfect-match (PM) probe signal intensity. The generated matrix was subsequently used for an analysis of *B. oleracea* response to a mineral nutrient stress. In another study Affymetrix Arabidopsis ATH1 were used for analyses of tropic responses (Esmon et al. 2006). The results suggested that the activation and accumulation of specific genes correlates with a lateral gradient of auxin in *Brassica* hypocotyls. *B. oleracea* anther-specific mRNAs were analyzed by using *Arabidopsis* cDNA macroarrays (Amagai et al. 2003). Similar *Arabidopsis* cDNA macroarrays have been used for a comparative analysis of drought stress response in *A. thaliana* and *Brassica oleracea* (Fig. 9-1) (Ludwików A., Babula-Skowronska, Szczepaniak M., Belter N., Dominiak E., Nowak W., Sadowski J., unpubl.). All these studies support the idea that *Arabidopsis* DNA microarrays and macroarrays with a high degree of confidence can be used to address gene expression profiling in *Brassica* species.

![Figure 9-1 Macroarrays for the analysis of Brassica oleracea spp. alboalbura line A12DH gene expression in response to drought stress. The figure shows a representative macroarray hybridization result for control (A) and drought-treated (B) *B. oleracea* samples.](image)
Due to the remarkable nucleotide sequence similarity between *B. oleracea* and *A. thaliana* (reported as high as 80%), full-length cDNA clones and EST resources for *A. thaliana* and other species of Brassicaceae are the most relevant for the precise annotation of new genes. To improve *A. thaliana* and *B. oleracea* annotation approximately 600,000 *Brassica oleracea* shotgun reads were sequenced by TIGR/CSHL/Washington University and analyzed by using the *Arabidopsis* genome and proteome databases (Katari et al. 2005). The authors focused their attention on the short arm of chromosome 4 of *A. thaliana* to test whether putative genes conserved between *Brassica* and *Arabidopsis* were expressed in comparison to non-conserved putative genes. The results revealed that putative genes tend to be expressed if they are identified in conserved regions. Additionally, the authors improved the *Arabidopsis* annotation by identifying new genes, demonstrating the expression of several hypothetical genes, and extending the *Arabidopsis* transcriptome by 850 genes. Moreover, incorrect gene structure was identified in approximately 21% of hypothetical genes, demonstrating the need of a re-annotation of this part of the *Arabidopsis* genome. Another comparative approach (Ayele et al. 2005) used the genome-wide alignment of Whole Genome Shotgun (WGS) sequences from *B. oleracea* to identify poorly annotated genes and genomic regions. The analysis resulted in the identification of conserved *Arabidopsis* genome sequences (CAGSs) that were subsequently merged into a potential functional unit. Moreover, several novel conserved units were identified that did exist apart from gene annotations. Overall, these studies demonstrate how information from *Brassica oleracea* sequences can be utilized to improve the annotation of the *Arabidopsis* genome.

Comparative genomics makes it possible to discover a large variety of functional elements by their conservation between *Arabidopsis* and *Brassica* (Ruvinsky and Ruvkun 2003; Doniger et al. 2005). Haberer et al. (2006) have used this strategy to study sequence conservation in orthologous promoter pairs. The employed methodology resulted in the identification of a nucleotide pattern in co-expression groups in *Arabidopsis* as well as putative orthologous upstream sequences in *Brassica*. Among many newly identified motifs, several known sequence patterns were revealed, which validated the undertaken approach. In another example, an *Arabidopsis* ATH1 was successfully used for the analysis of the flower (anther) development in the *Brassica oleracea* Ms-cd1 mutant (Kang et al. 2008). In this particular approach, microarray profiling facilitated the identification of differential expression between male sterile Ms-cd1 and wild type (male fertile) plants. Importantly, kinases, cell wall modification genes, and ion transport genes were suppressed by Ms-cd1. The authors concluded that Ms-cd1 might affect an anther developmental pathway, which influences genes involved in signaling, cell wall modification, and ion transport in pollen grains.
9.3.2 Arabidopsis thaliana—Brassica rapa

Recent information obtained from the *A. thaliana* model has advanced our understanding of defense responses. However, there is still limited knowledge about the pathogen defense mechanism in related *Brassica* species. It is likely that related species may have unique pathogen responsive mechanisms and regulation of gene expression, which does not correspond to well known plant models. To investigate the above assumption the complexity of the gene expression of *A. thaliana* and *B. rapa* were examined in plants infected with fungal pathogen *Colletotrichum higginsianum* as well as after salicylic acid (SA), ethylene (ET), and methyl jasmonate (MeJA) treatment (Narusaka et al. 2006). Transcriptome analysis was performed using a cDNA microarray with 1,820 cDNA clones selected from the cDNA library of Chinese cabbage and 1,2K *Arabidopsis* cDNA microarray. In general, the constructed *B. rapa* microarray plate represented putative defense-related and regulatory genes. Most gene expression profiles between *B. rapa* and *Arabidopsis* were similar suggesting conservation of these genes function. However, genes that undergo different expression patterns between *B. rapa* and *Arabidopsis* were also identified revealing minor differences in *B. rapa* response to fungal pathogens in comparison to *A. thaliana*.

9.4 Application of Functional Genomics in Genomics-assisted Breeding

In the past decade, plant breeders have made a huge effort to create germplasm and plant cultivars having desirable traits. From these studies we have learnt that the success of this selection relies on the availability of wild alleles for the agronomic traits. As in many other branches of life science, plant breeding for crop improvement cannot exist without transcriptomics. By applying functional genomics approaches we are now capable of providing candidate genes that can be simultaneously used for understanding the biology of the plants and development of agronomically important traits (Ma et al. 2004; Tuberosa and Salvi 2006; Heffner et al. 2009; Sarosh et al. 2009). In general, the transcriptomics can be used for identification mechanisms that control the biological system. This initial knowledge serves as a basis for further molecular analyses, including the generation of transgenic plants and time-consuming studies of emerging phenotypes. It is important to note that despite such a huge potential of transcriptomics and other genomics methods there are still many bottlenecks. One of them is the variability of the design phenotype in a field. Simply, many transgenic plants do not copy its *in vitro* or chamber performance in the field. Nevertheless, it is highly desirable if the quality of the improved crop meets both yield and consumer satisfaction.
Despite all the undergoing breeding programs and the recent technological breakthroughs the overall contribution of genomics-assisted breeding to the release of new vegetable *Brassica* cultivars is almost none. So far we can point out just a few studies that have put new focus on breeding strategies for seed improvement and stress tolerance. These strategies utilize mainly classical techniques such as genetic mapping, association mapping, QTL mapping and traits combination (Tuberosa et al. 2002; Lou et al. 2007; Zhao et al. 2008; Heffner et al. 2009), and these will not be discussed here. Fortunately, genomics-based approaches for phenotype predictions utilizing microarrays as a tool are also accumulating. DNA microarrays (oligonucleotide and cDNA micro- and macroarrays) have been utilized to identify differentially expressed genes in a few *Brassica* species, including *B. rapa* and *B. oleracea*. Reported studies have made an initial but important contribution to vegetable *Brassica* crop biology. Applying gene expression profiling is now possible to identify regulatory gene networks that can be useful for trait-specific screens of agronomically important alleles.

### 9.4.1 Seed Development and Quality

Seed development is an important stage for seed quality (Wszelaki et al. 2007). Thus the understanding of the mechanisms controlling seed development and germination is essential to achieve high harvest yields (Soeda et al. 2005). Recent transcriptome analyses have advanced our knowledge of seed processing by providing information on how multiple genes interact to control the effect of priming (Soeda et al. 2005; Hudson et al. 2007; Xiang et al. 2008). From these analyses, it is now possible to observe how priming improves seedling performance and how can it be used in the field.

Two groups have successfully applied microarrays to study the effect of priming. Hudson et al. (2007) used Affymetrix ATH1 array to determine the mechanism of priming on the germination efficiency of cauliflower seeds (cultivars Lintop and Maverick). Microarray analyses differentiated three distinct expression patterns that correlated with hydration treatment. Additionally, several functional categories of genes were identified to have particular expression profiles in investigated conditions. The authors propose that the effect of the priming process on the induction of protein synthesis could be a factor that determines seed durability. Furthermore, among 833 statistically significant *Arabidopsis* homologs, 59 were classified to be involved in protein turnover, protein fate, and ribosome biogenesis. Moreover, the majority of the genes were identified as ribosomal proteins, or were related to global metabolism. Significantly, the validation of results using real-time quantitative PCR confirmed the expression of four out of five
orthologous *Brassica* transcripts, indicating slightly lower reproducibility of microarray hybridization results.

A second group (Soeda et al. 2005) investigated priming and germination exploiting cDNA microarrays containing genes that were isolated from cotyledon or germinating *B. napus* seeds. The experiment design, in comparison to the previous study, was supposed to prevent information loss occurring in comparative approaches. Significantly, applying strict criteria for detection of hybridization signal this group selected 1,100 cDNAs corresponding to 850 distinct *B. napus* genes. In total, clustering the samples revealed five significant expression patterns that depend on preferential expression in dry, osmoprimed or germinating seeds. Although, some parts of the results were rather confirmatory, a vast amount of data provided additional information of the gene expression levels in relation to the seed germination, osmopriming, and stress tolerance. Notably, the authors showed that the expression of stress-related genes predominantly occur during seed maturation and slow drying. Most of these genes were known to encode LEA proteins or ABA-inducible dehydrins (RAB18, Em6). In addition, putative signaling pathways involved in the regulation of these genes have also been revealed based on an analysis of data deposited in Genevestigator. Overall, these data suggested that the regulation of osmopriming is complex and to some extent overlaps with germination pathway. Therefore, identification of putative regulators is important for the understanding of germination programs and seedling survival in stress conditions. It is important to note that an insight gained from this study provided new gene candidates for seed quality.

Another useful approach to gene function study was applied by Xiang et al. (2008). To test the quality of the seed cDNA microarray they analyzed different stages of *Brassica* seed development and different types of tissues. The authors identified differentially regulated genes between embryo and leaf that are consistent with other findings in the field. Further functional analyses of these genes should provide a rich source of information on the biology and technology of the seeds.

**9.4.2 Tolerance to Abiotic Stresses**

Plants are continuously exposed to abiotic stresses and have evolved mechanisms to reduce their effects. Various abiotic stresses limit vegetable *Brassica* production and the most important include water stress (Lee et al. 2008; Ashraf et al. 2001; Ludwików unpubl.) (Lee et al. 2008; Ashraf et al. 2001; Ludwikow unpublished), nutrient deficit (poor degraded soil) (Hammond et al. 2005), high temperature (Björkman and Pearson 1998), wounding and pathogen attack (Park et al. 2005; Broekgaarden et al. 2007; Bruinsma et al. 2007; Sarosh et al. 2009) (Broekgaarden et al. 2007;
Broekgaarden et al. 2008; Bruinsma et al. 2007; Park et al. 2005; Sarosh et al. 2009). However, with the exception of biotic stresses, which will be discussed later, there are only two studies reporting the profile of Brassica plants in abiotic stress conditions (Hammond et al. 2005; Zhou et al. 2007; Lee et al. 2008). Gene expression profiling are novel approaches for the determination of genes and pathways involved in stress response and tolerance. Therefore, we are still far from detailed knowledge of signal transduction cascades activated under abiotic stress response in Brassica plants. Nevertheless, based on these initial reports, we summarized the progress and the current status of the gene profiling in the studies of Brassica stress response. We also discussed the contribution of these studies to Brassica breeding. Particular emphasis has been made to study B. oleracea transcriptional response to phosphorus stress (Hammond et al. 2005). The reported study has been designed to improve our knowledge about the regulation of gene expression in mineral nutrient stress and to investigate the consequences of P starvation. The analysis utilizing the Affymetrix ATH1 Genechip array revealed approximately 100 genes identified as significantly differentially regulated in the shoots in the response to phosphorus starvation. Only a few functional categories of genes were identified to have distinct expression profiles in P stress conditions. Among them several genes involved in the P turnover during P starvation were found to be up-regulated. Notably, many transcripts or functional groups identified for P starvation have been previously described in A. thaliana. Nevertheless, many identified genes can serve as good candidates for genetic manipulations improving nutrient efficiency in plants. Another important study, developed to identify defensive responses against UV light in turnip (Zhou et al. 2007), indicated distinct genes regulated by UV-B and UV-A treatments while similar sets of genes were regulated by blue and UV-A light treatment. Several functional categories were identified to have particular expression profiles in stress conditions. As expected, ethylene- and oxidative stress-related genes showed higher expression levels upon UV-B exposure, while lower expression were observed for tissue or cell growth-related genes. In addition, what was the most important, a group of genes related to anthocyanin biosynthesis were found up-regulated in response to UV-A and this correlated with higher anthocyanin level in treated plants. Notably, this study has revealed a putative anthocyanin biosynthesis pathway mediated by unknown UV-A-specific photoreceptor. Anthocyanin pigment content has a critical role in the color quality of many fruits and vegetables, simply by making them attractive to the consumer. Thus, these results have great potential for practical applications in prevention and quality control in crop production.
As for many other crop species, the advances made in genomic analysis were limited by the extremely slow development of genomic resources including the construction of Brassica microarray. The most recent transcriptomic experiment using Brassica 24K Oligo Microarray (KBGP-24K) was designed to identify cold, salt and drought stress-responsive genes in B. rapa (Lee et al. 2008). By a cross-species comparison the authors selected many genes specifically responding to each factor: cold, salt, and drought, as well as a group of generally responsive genes regulated by all three forms of stress. Significantly, genes encoding several novel transcripts were identified and the differential expression of many known stress-induced genes was confirmed. This includes an ortholog of Arabidopsis CRT/DRE-binding factor gene (BnCBF17). Additionally, genes encoding putative transcription factors (members of AP2-EREBP, MYB, bHLH, bZIP, homeobox, and heat stress transcription factor families) responsible for abiotic stresses were also recognized. Taken together, the expression patterns for the majority of differentially expressed genes were rather similar to Arabidopsis homologues. The exception is B. rapa SOS genes whose expression level was not significant in this study. Nevertheless, there is a notable outcome of this research showing that such a direction is worth continuing. Identified candidate genes should be further evaluated for their phenotype in the field.

9.4.3 Transcriptomics to Improve Disease Resistance

Little is known about pathogen defense mechanisms in Brassica species. Although vegetable Brassica crop losses due to invading pathogens such as insect pests and fungi are significant, progress on identifying defense mechanisms that govern the response is extremely slow. It has been established that plants defend themselves against pathogens by activating both direct and indirect defense mechanisms (van Poecke and Dicke 2002; Ellis and Dodds 2003; Broekgaarden et al. 2007) (van Poecke and Dicke, 2002; Ellis and Dodds, 2003; Broekgaarden et al. 2007). In addition, there are several major signal transduction pathways involved in plant defense including salicylic acid (SA), jasmonic acid (JA), ethylene, ABA and the reactive oxygen species pathways (Fig. 9-2). However, our understanding of how these components interact and how other components are involved is still limited. Studies in Arabidopsis have demonstrated that an SA-dependent response is deployed against biotrophic pathogens, whereas ethylene or JA responses are important for the induction of resistance to insect and arthropod herbivores, some necrotrophic fungal pathogens, and nonpathogenic fungi (Thaler et al. 2004; Toth and Birch 2005; Park et al. 2005; Toth et al. 2006; Narusaka et al. 2006). Available gene-expression profiling studies analyzing host-pathogen interactions in Brassica (Park et al. 2005;
Ferry et al. (2006); Tuberosa and Salvi (2006); Broekgaarden et al. (2007); Bruinsma et al. (2007) have not revealed additional members of defense signaling pathways, but signify an interaction between JA, ET and SA signaling as crucial components of defense important for the understanding of plant response to biotic stresses.

Figure 9.2 Signaling network in plant defense against pathogens.

To identify and characterize genes that are involved in inducible direct defense in *Brassica*, Broekgaarden et al. (2007) analyzed the gene expression profiles of two *B.oleracea* cultivars, Rivera and Christmas Drumhead, upon feeding by *Pieris rapae*. The two cultivars were shown to differ in transcriptional responses and the level of direct defense against pathogens. The analysis revealed a higher direct defense in Rivera compared to Christmas Drumhead cultivar. This correlated with differences in the transcriptional response including timing and cultivar-specific activation of some genes. However, what was most important, this study indicated that the jasmonate signaling network did not control a large portion of *P. rapae*-inducible genes. These results demonstrated that direct defense against *P. rapae* was also influenced by other undefined factors.

Nevertheless, the involvement of JA in the regulation of direct and indirect defense has already been documented in many plant species (Zhu-Salzman et al. 2004; Liechti and Farmer 2006). JA is involved in the defense responses to oviposition and feeding damage (Hilker et al. 2005; Schroder et al. 2007; Walling 2008) (Schroder et al. 2007; Hilker et al. 2005; Walling, 2008). Several studies focused on the retardation of growth and leaf consumption of herbivores (Kingsolver and Woods 1997; Dunning et al. 2003; Williams et al. 2004) (Williams et al., 2004; Dunning et al., 2003; Kingsolver and Woods, 1997). In addition to these, Bruinsma et al. (2007) tested the effect of JA on the oviposition of *Pieris rapae* and *P. brassicae* herbivores in cabbage. Interestingly, they found that JA treatment reduces the acceptance of leaves for oviposition. However, JA-induced signaling
changed the level of the insect’s tolerance of leaves. Another study aimed at the characterization of the gene expression profiles in *B. rapa* in order to investigate the key lines of defense against *C. higginsianum* infection (Narusaka et al. 2006). The results confirmed that SA, JA and ET-dependent defenses are essential for host-pathogen interaction, however ET-dependent signaling pathway may be a key player in the early defense. Similarly, a cDNA microarray was utilized in the analysis of the defence response in canola upon treatment of the fungal pathogen *Sclerotinia sclerotiorum* (Yang et al. 2007). This analysis led to the identification of 300 transcripts that were responsive to this pathogen-challenge. In agreement to previous studies, this analysis also identified genes that were associated with JA biosynthesis and signaling. It is important to note that among identified genes an up-regulation of genes orthologous to AtERF2 and AtERF4 were reported, further confirming the involvement of ERFs in the defense pathways in related plant species. Additionally, the importance of the salicylate-dependent and ethylene-mediated pathways in defense responses to *Pseudomonas syringae pv. tomato* in Chinese cabbage was confirmed in the study of Park et al. (2005). Overall, nearly all of the presented results are in agreement with previous *Arabidopsis* studies. Again, further functional characterization of the revealed members of the defence signaling might be crucial for future rapid *Brassica* yield improvement.

9.5 Conclusions

Transcriptome analyses are powerful approaches for the decoding of gene functions. Due to a remarkable technical advance in transcriptomics new genomic resources have been developed for *Brassica*, an agriculturally important genus consisting of several vegetable crops. Many ESTs and cDNA libraries that have been generated were subsequently used to design *Brassica* DNA microarrays. As expected available microarray studies provide evidence for the effective application of microarrays to identify a large number of genes involved in seed growth, development, disease resistance, and abiotic stress tolerance. Despite the substantial progress that has been made toward uncovering the network of gene functions in investigated conditions, much work still needs to be done to completely characterize associations between gene expression patterns and the genotypes in breeding populations. Although functional genomics tools are very helpful in the detailed molecular description of agronomic traits, they alone cannot improve our understanding of biological systems shaping the phenotype. Therefore, much more attention should be paid now to the integration of high-throughput genomics with routine breeding strategies. The question remains: how will we take advantage of these opportunities in the future?
Acknowledgements
This study was supported by the Polish Committee for Scientific Research Grants PBZ-KBN-089/P06/2003, MNI 2P04A03329, PBZ-MNiI-2/1/2005, PZ-19/2006/2/JS, PBZ-MNiSW-2/3/2006/19/IGR/4 and 682/N-COST/2010/0. I. Makalowska is acknowledged for critical reading of the manuscript.

References


ABSTRACT

Proteome and metabolome are very complex and dynamic as they differ in their composition depending on many factors such as cell and tissue type, stage of development, and environmental conditions. Though proteome and metabolome deliver a large amount of data, there are not many reports describing whole proteome or metabolome. Not only are they complex, but there are also difficulties in reproducing their analyses. As plants of Brassicaceae family deliver food that is important for human nutrition and health, a large part of the conducted experiments concern the amount and composition of metabolites (especially glucosinolates) or proteins involved in anticancer activity. In recent years, there has been significant development of proteomic techniques, especially those related to mass spectrometry. In this chapter, we focus on the methods and techniques which have been most frequently used in research on proteome and metabolome in Brassica species.

Keywords: Brassica proteome, Brassica metabolome, proteome profiles, stress response, glucosinolates, anticancer activity

10.1 Introduction

The role of large scale approaches in transcriptome, proteome and metabolome analysis applied in gene expression studies is continuously growing. Numerous reports on selected individual transcripts, proteins, and metabolites could not resolve the problems connected with plant responses to, for example, environmental cues or gene manipulations performed by breeders and geneticists. Dynamic development of transcriptome profiling tools enabled deeper insight into many cell pathways and gene
co-regulations over the past decade (see Chapter 9 for details). However, whole transcriptome analysis is just the first step in the study of gene expression because the correlation between changes in mRNA and protein content is not linear (Kersten et al. 2002; Millar 2004).

The large scale gene expression analyses, conducted so far, have concentrated mostly on model plants. Thus, a lot of information concerning the fundamental aspects of plant biology is available, although it needs to be emphasized that the course of some processes can be different in model plants and in crop plants. Vegetables such as cabbage, cauliflower, broccoli and others, belonging to the Brassicaceae family, are known to be important for human nutrition. Some unique metabolites produced in these plants have antitumor activity. In large scale analysis for better understanding of the network of connections between proteins and metabolites, it is of great importance to analyze simultaneously the proteins and metabolites, which would permit obtaining plants more resistant to stress and containing more compounds beneficial for human health.

On the other hand, analysis of non-model organisms is rather challenging because of the lack of coding sequence information, which makes protein identification more complex. Some strategies have been developed to circumvent these limitations. In the first strategy, peptide sequences are obtained in the mass spectrometry study. At present, several tools for bioinformatic peptide sequencing are available (for example NovoHMM and PepNovo algorithms). In the second, the data obtained are compared to the protein database of related organisms. This method allows identification of only those proteins that have at least one peptide with a sequence identical to the peptide from a protein in the database. In terms of the Brassicaceae species, such a comparative sequence analysis is facilitated by the fact that they are closely related to the model plant Arabidopsis thaliana, which is a member of the same family (van Wijk 2001; Grossman et al. 2007). This coding sequence homology enabled extensive comparative studies at the gene as well as protein levels within Brassicaceae species. For example, the antibodies specific for four A. thaliana superoxide dismutases (Kliebenstein et al. 1998) recognize their homologs from different species of Brassica genus (Fig. 10-1; Misztal et al. unpubl.).

Another field of large scale analyses, though poorly established as yet, is metabolite profiling. Metabolites are the end products of cellular processes. Their level and composition differ depending on the developmental level, tissue or in response to environmental changes. The analyses reported so far concerned the quantification of a few defined metabolites or identification of the overall metabolic changes. None of these analyses provides explanation of how metabolic networks work. The majority of recent studies concern the identification and quantification of multiple targets. Large scale analyses
require reliable and high-throughput methods of isolation and analysis of proteins and metabolites.

Figure 10-1 Specific identification of homologous variants of two superoxide dismutases Western blot detection of two superoxide dismutases in protein extract from *A. thaliana* and three members of Brassicaceae family based on antibodies raised for *A. thaliana* isozymes (L. Miszta et al. unpublished). 1-*A. thaliana*, 2-*B. oleracea* var. *alboglabra*, 3-*B. napus*, 4-*B. rapa* ssp. *pekinensis*. Antibodies used were kindly provided by Prof. Daniel J. Kliebenstein.

10.2 Proteomics and Metabolomics Analysis

In this chapter, we focus on the presentation of the methods and techniques most frequently used in proteomic and metabolomic research. The development of transcriptomics, proteomics and metabolomics-oriented approaches, tools (including bioinformatics and statistics), technologies and equipment granted specialized “omics” platforms installation in dozens of research centers. They allow detection and throughout analysis of many molecules simultaneously.

In proteomics, two techniques play important roles: two dimensional electrophoresis (2-DE) for protein separation and mass spectrometry (MS) for protein sequencing and identification. Additionally, in proteome and metabolome studies mass spectrometry analyses are coupled with liquid chromatography (LC), which is known as the LC/MS technique allowing precise separation of different kinds of molecules. The general procedure for proteome analysis includes four steps: (1) sample preparation (extraction, digestion and enrichment), (2) separation of the proteins or the digested peptides, (3) identification of the proteins by mass spectrometry (top-down or bottom-up strategy) and (4) data mining for the information generated (functional classification or comparative analysis) (Fig. 10-2; Aebersold and Mann 2003).

10.2.1 Sample Preparation for Proteomics or Metabolomics Studies

For biologists the most important part is the correct sample preparation that can be followed by precise analysis, most commonly, by mass spectrometry. The choice of an efficient method for protein or metabolite extraction from cells or tissues and then enrichment of fractions of interest appears to be crucial in obtaining reproducible data. Proteomes and metabolomes are
very dynamic and could very quickly change their composition because of a variety of cell and tissue types that are analyzed and the complexity of samples prepared.

For this reason, the proteomes or metabolomes are unique for each tissue, cell or organelle that unfortunately can change very quickly under the influence of different stimuli. This changeability is still a challenge in the precise preparation of samples for plant proteomic and metabolomic experiments. In contrast to DNA extraction techniques there are no standard procedures for protein extraction from different plant cells, tissues, or organelles; thus, specific extraction methods must be developed. Additionally, the presence of cell walls in plant material makes the protein extraction procedure more complicated in plant than in animal studies. This difference requires additional steps in plant extraction procedures (James 2004). The first step of all extraction methods for metabolomic and proteomic analyses must prevent sample degradation and rapidly stop all cellular enzymatic activities and metabolic processes that could change the metabolite or protein contents. The relatively high level of different secondary metabolites present in plant tissues could be problematic for accurate proteome analysis. This is important especially for successful high-quality 2-DE separation. Thus protein samples must be also sufficiently purified from salts, surfactants or other biomolecules (lipids, carbohydrates). Plant tissues usually contain a high concentration of proteases, therefore the extraction buffers used should also include inhibitors of proteases. For more information about sample preparation for proteomic and metabolomic analysis readers are directed to several reviews (Herbert et al. 2007; Hurkman and Tanaka 2007; Thelen 2007; Wang et al. 2008).
10.2.2 Molecules Separation in Proteomics and Metabolomics

Currently in proteomics, there are two core techniques for separation of complex protein samples: electrophoresis and chromatography. Among several electrophoretic methods the most popular are: denaturing polyacrylamide gel electrophoresis (SDS-PAGE) and two dimensional electrophoresis. Additionally, chromatographic techniques, especially liquid chromatography are also used extensively for protein purification and separation of peptides and small molecules (Table 10-1).

10.2.3 Identification of Molecules (Mass Spectrometry, Nuclear Magnetic Resonance)

The separation of hundreds of proteins by 2-DE was already possible in the late 1980s, but for proteomics researchers the lack of useful tools for rapid and robust identification of proteins of interest were still problematic. Before mass spectrometry, the most common approach to identify 2-DE separated proteins was the method called Edman degradation, based on enzymatic digestion of amino terminus of intact proteins. Released amino acids were identified by UV-absorbance spectroscopy. This technique was well established and automated but was also slow and had relatively poor sensitivity (Patterson and Aebershold 2003; Domon and Aebershold 2006). Currently, the most commonly used methods for molecules identification are mass spectrometry and nuclear magnetic resonance spectroscopy (NMR) (Table 10-2).

Mass spectrometry is a highly sensitive and accurate method for determination of molecular mass of different molecules. This method is especially important in proteomic approaches because it is unique and gives an opportunity to determine the protein sequence from very complex protein or peptide mixture. The first report on mass spectrometry was published at the beginning of the last century by J. J. Thomson (1913). Since that time MS has become a well established technique for analysis of chemical structure of small molecules. MS is based on molecule ionization. For analysis of high-mass biological samples (peptides or proteins) researchers had to wait for decades because of the lack of suitable soft-ionization techniques for biological samples (which allows one to ionize peptides without the breakdown of peptide bonds). Two soft desorption ionization methods for mass spectrometry of biological macromolecules were discovered in the late 1980s: matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) (Tanaka et al. 1987; Karas and Hillenkamp 1988; Fenn et al. 1989). For this discovery, John Fenn and Koichi Tanaka were awarded one-quarter of the Nobel Prize for chemistry in 2002.
Table 10-1 The techniques most commonly used for proteins, peptides or metabolites separation.

<table>
<thead>
<tr>
<th>ELECTROPHORETIC METHODS</th>
<th>Technique</th>
<th>Principle of sample separation</th>
<th>Advantages/Disadvantages</th>
<th>Type of molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>One dimensional electrophoresis (SDS-PAGE)</td>
<td>• proteins are separated on denaturing (SDS) polyacrylamide gel according to molecular mass</td>
<td>• for protein separation only&lt;br&gt;• relatively low resolution&lt;br&gt;• separation acc. to molecular mass only</td>
<td>Proteins</td>
<td></td>
</tr>
<tr>
<td>Two dimensional electrophoresis, (2-DE)</td>
<td>• first dimension: isoelectrofocusing (IEF)—proteins are separated acc. to isoelectric point second dimension: SDS-PAGE</td>
<td>• the first method for high-throughput proteins separation&lt;br&gt;• high resolution of protein separation</td>
<td>Proteins</td>
<td></td>
</tr>
<tr>
<td>2-D difference gel electrophoresis (2D-DIGE)</td>
<td>• protein samples are labeled with different fluorescent dyes and then separated by 2-DE (in the same gel)&lt;br&gt;• gel is scanned by a laser scanner</td>
<td>• internal standard&lt;br&gt;• useful for comparative proteomics&lt;br&gt;• high resolution of protein separation&lt;br&gt;• fluorescent dyes are more sensitive than silver nitrate staining</td>
<td>Proteins</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHROMATOGRAPHIC METHODS</th>
<th>Technique</th>
<th>Principle of sample separation</th>
<th>Advantages/Disadvantages</th>
<th>Type of molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column chromatography</td>
<td>• the sample is injected into mobile phase (liquid or gas). The mobile phase with the sample moves through stationary phase (immobilized porous substance)&lt;br&gt;• individual components have different retention times (RT) due to the differences in their interactions with phases&lt;br&gt;• in normal phase (NP) chromatography the mobile phase is less polar than the stationary phase In reverse phase (RP) system is in opposite</td>
<td>• molecules can be separated acc. to molecular weight, charge, hydrophobicity or affinity&lt;br&gt;• liquid chromatography, (LC) or gas chromatography (GC) are most commonly used for proteomics and metabolomics</td>
<td>Proteins, peptides, other small molecules</td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| High pressure liquid chromatography  | - high pressure pumps are used to get a more resonate flow rate and enhance the separation efficiency  
| (HPLC)                               | - separation efficiency depends mainly on the balance of the affinity of the sample components to the stationary phase and their solubility in mobile phase  
|                                      | - currently most widely used for proteomics and metabolomics  
|                                      | - the eluate may be applied directly into a mass spectrometer by electrospray ionization (on-line) or collected for subsequent analysis (off-line)  
|                                      | - good efficiency and resolution of separation  
| Reverse phase chromatography (RP-HPLC) | - the stationary phase is less polar than the mobile phase  
|                                      | - molecules are separated on the basis of differences in hydrophobicity  
|                                      | - the stationary phase usually prepared by surface-modification of silica beads (modifications cause the absorbance of the sample components)  
|                                      | - currently most widely used for proteomics and metabolomics  
|                                      | - the eluate may be applied directly into a mass spectrometer by electrospray ionization (on-line) or collected for subsequent analysis (off-line)  
|                                      | - good efficiency and resolution of separation  
| Strong cation exchange (SCX)         | - the stationary phase is usually made of surface modified sulfonic acid group  
|                                      | - two opposite charges attract each other  
|                                      | - SCX is one of four types of ion exchange chromatography (the others are: weak cation exchange (WCX), strong anion exchange (SAX) and weak anion exchange (WAX))  
|                                      | - after SCX the mobile phase contains salts and eluate can’t be directly applied to the mass spectrometer  
|                                      | - good efficiency and resolution of separation  
| Affinity chromatography (AC)         | - technique based on specific interaction between separated analyte and specific functional group on stationary phase  
|                                      | - the components of interest are retained while non interacting components pass through the column  
|                                      | - Proteins, peptides  
|                                      | - Proteins, peptides, other small molecules  

### Table 10-2 The techniques most commonly used for proteins, peptides or metabolites identification.

**MASS SPECTROMETRY**

<table>
<thead>
<tr>
<th>Technique</th>
<th>Principle of sample analysis</th>
<th>Advantages/Disadvantages</th>
<th>Type of molecules</th>
</tr>
</thead>
</table>
| Mass spectrometry (MS)         | • first step: molecules are ionized (e.g., MALDI, ESI)  
• second step: ionized molecules are separated acc. to their mass-to-charge ratio (m/z) with different types of analyzers (e.g., TOF)  
• third step: ions detections  
• fourth step: *in silico* protein identification (with algorithms)  
• highly sensitive and accurate method for the determination of molecular mass for different types of molecules | Protein, peptide, small molecules                                                      |
| Tandem mass spectrometry (MS/MS) | • peptide ion should be intact (peptide bonds) until fragmentation step  
• peptide are ionized with soft ionization techniques  
• two mass analyzers connected in tandem (e.g. TOF/TOF); the first mass analyzer characterizes the m/z of intact peptides; subsequently, selected peptide is fragmented and the other mass analyzer characterizes the m/z of ions  
• for protein (peptide) sequencing commonly used system is tandem mass spectrometry coupled which HPLC technique (HPLC-MS/MS)  
• determination of the structure of small molecules  
• high resolution and efficiency  
• requires low amount of a sample | Protein, peptide, small molecules                                                      |

**NUCLEAR MAGNETIC RESONANCE**

<table>
<thead>
<tr>
<th>Technique</th>
<th>Principle of sample analysis</th>
<th>Advantages/Disadvantages</th>
<th>Type of molecules</th>
</tr>
</thead>
</table>
| Nuclear magnetic resonance (NMR) | • isotopes with the spin and magnetic moment yields nuclear magnetic resonance (NMR) and can be observed in a strong magnetic field pulsed with electromagnetic radiation  
• the majority of biologically important elements occurs in isotopic forms. Some of these have integral (N^{14}) or fractional spin (C^{13}, H^{1}, N^{15}, F^{19}, P^{31}) and few have no spin (C^{12}, O^{16}) | provides detailed information about topology, dynamics and three-dimensional structure of metabolites or proteins | Protein, small molecules, metabolites  |
These ionization techniques together with improvement in sample protein preparation methods and the increasing number of full genome sequences (high quality protein databases) were crucial for high-throughput protein identification by mass spectrometry techniques. Currently, different kinds of spectrometers are used for measuring the masses of molecules (for review see Smith 2002; Domon and Aebersold 2006). Presently, the most commonly used is tandem mass spectrometry (MS/MS) coupled with high-pressure liquid chromatography (HPLC) technique. MS/MS analysis generates data on molecular mass and sequence of proteins. Due to MS development two major approaches are used for protein identification and characterization: the top-down and bottom-up approaches (also called shotgun proteomics or peptidomics). In the top-down approach the intact proteins are directly used for analysis and in the bottom-up approach the proteins are previously digested with an efficient protease (mostly trypsin) and peptides obtained are then used for analysis. All common mass spectrometers consist of three functional units: 1) ion source 2) mass analyzer and 3) ion detector and data acquisition unit (Eidhammer et al. 2007).

10.3 Current Status of Vegetable Brassica Proteomics

Obviously, most of the proteomics research within the Brassicaceae family concerns the model plant *A. thaliana* while there is not much information on the proteome of vegetable Brassicas. A few available publications describe large scale analyses on proteome comparison to identify genetic relationships within the Brassicaceae family (Marquez et al. 2001), pollen proteome (Murphy 2006), autopolyploidy relations in cabbage (Albertin et al. 2005) or pathogen attack response (Subramanian et al. 2005; Widarto et al. 2006). Besides general proteomic analysis, the majority of works concern individual proteins. Typically, analysis of a whole proteome focused on selected, well defined stress conditions, tissue, organ, organellum or developmental stage. Interestingly, it is not only the cost that limits wide scale comparative study. Complexity of the proteome system under study requires concentration on simple models. In addition, comparative studies of proteomes at different developmental stages or several time points during exposition to stress seem to be not sufficiently reproducible and challengeable at this stage of proteomics development. The proteome analysis can be limited for proteins of low abundance. These proteins can be masked in the presence of very abundant ones. Another problem concerns the proteins of such amino acid sequence that hampers their isolation or identification because of the problems with dissolving, digestion or ionization. Analysis of such proteins could be facilitated when a study concerns only one or a few proteins. Another advantage of analysis of a few proteins is the possibility of tracing the qualitative and quantitative
differences in their expression in several developmental stages, tissues and stresses.

10.3.1 Structural and Developmental Proteome Profiles

Generally, a large part of proteomics-based research concerns the analysis of organellar proteome. In plants, proteome research covers most frequently chloroplast and mitochondrial proteomes. Pawlowski et al. (2005) described soluble and membrane bound fractions of cauliflower mitochondria with the use of two dimensional electrophoresis. The patterns obtained were evaluated with Imager Master 2-D Elite software. After silver nitrate staining, 561 protein spots were detected. The study of mitochondrial proteome conducted by Hochholdinger et al. (2004) showed that mitochondrial proteins encoded by the nuclear genome could differ even if nuclear genomes were identical. It suggests that mitochondrial genotypes can regulate the accumulation of such proteins. In the research conducted by Pawlowski et al. (2005) mitochondria were isolated from cauliflower buds by differential centrifugation in Percoll gradient. After fractionation into soluble, integral membrane and peripheral membrane, the mitochondrial proteins were resolved by 2-D electrophoresis. The patterns obtained were compared to the A. thaliana mitochondrial proteome. About 30 proteins showed similar pI and MW to the A. thaliana proteins, so they could be putative orthologs. They were for example pyruvate dehydrogenase, prohibitin or the elongation factor TU. Some of cauliflower proteins probably have different hydrophobicity and abundance than those of A. thaliana proteins, which makes their identification difficult. Some of the protein spots were extracted from gel and analyzed by ESI/MS, e.g., adenylate kinase A-B, ATP synthase, etc. were identified in this way. Differences observed between subfractions were qualitative (for example nine proteins from membrane fraction were not observed in soluble fraction) and quantitative (39 proteins being about 1.5-fold more intense in the membrane fraction; 62 proteins more intense in the soluble fraction).

Schmidt et al. (2007) analyzed vacuoles isolated from cauliflower buds. In meristematic cells there are a great number of small vacuoles, while in the mature cells there is a large central vacuole. Thus far only a small number of vacuolar proteins have been identified. The authors isolated vacuoles from cauliflower buds, containing both meristematic and fully expanded cells. The isolated proteins were compared with A thaliana vacuolar proteome. The results allowed identification of 102 integral and 214 peripheral membrane proteins. The most prominent protein was phyrophosphatase. The authors also identified 57 proteins containing one or more membrane spanning domain. To verify the localization of the identified proteins, fusions of five of them with green fluorescent protein (GFP) were constructed. Four of these
proteins (viz., putative cation transporter, nodulin-21, senescence related membrane protein and one protein of unknown function) were localized in the vacuolar membrane, whereas the fifth (white-brown ATP-binding cassette transporter) was localized in the plasma membrane. When small vacuoles from young cells fuse to form a single large vacuole they must increase their volume. It is related to the transport of inorganic ions into the vacuoles followed by water influx. Vacuoles in mature cells control cell homeostasis by regulating the concentration of metabolites and ions in cytoplasm by their uptake, storage and release. So, proteins participating in the transport of these compounds are very important. The proton pumps generate an electrochemical gradient driving solute uptake; antiporters contribute to the accumulation of uncharged solutes as sucrose. Among the antiporters of great interest are the Ca/proton ones as their expression can lead to salt tolerance. Another interesting group is that of cadmium/proton antiporters as they could be used in phytoremediation. The authors isolated tonoplast proteins from cauliflower buds, determined their purity by Western blot with specific antibodies and analyzed the ATPase activity. They found that an isolated fraction contained up to 1.3% plasma membranes, chloroplasts, mitochondria and other cellular structures. Five different extraction procedures were used: alkaline, saline, acetone, chloroform/methanol extractions and treatment with sodium phosphate. Proteins were fractionated by SDS-PAGE. Before MS, tryptic peptides were separated by reverse phase chromatography (RP-HPLC). Proteins were identified by SEQUEST algorithm search against the *A. thaliana* database on National Center for Biotechnology Information (NCBI). Using the methods described, the authors identified 102 cauliflower membrane proteins. Fifty three proteins were identified by a unique peptide hit. There were 28 proteins in saline, 36 in acetone, 13 in alkaline and 27 in phosphate extracted fractions. Of the 102 proteins, 56 proteins (55%) were isolated solely via one of the applied procedures. Fifty seven membrane proteins were not localized, and 36 could be annotated as non-vacuolar proteins. Four proteins localized in vacuolar membrane by GFP fusions have an unknown function. One of them (homologue of At1g16390) could be a putative cation transporter. This protein has 11 transmembrane domains.

One of the most frequently analyzed tissues and mechanisms are pollen, stigma and their interactions. It has been known for a long time that some proteins are very important in mediating self-incompatibility. Protein/glycoprotein constituent of pistil specifically inhibits self pollen growth. In *Brassica oleracea* stigma S-allele-specific glycoproteins have been detected. S-specificity resides probably in the protein part of the glycoprotein molecule. It has been also shown that one glycoprotein causes the inhibition of pollen applied to its surface. However, the manner of this interaction/inhibition is still unknown. The analysis by Roberts et al (1984) concerns
pollen-stigma interactions and impact of movement of proteins. The authors observed that mature stigmas treated with cycloheximide lose their self-incompatibility. This means that for the control of self-pollination, synthesis of new proteins is necessary. On the other hand, tunicamycin application has no effect. Furthermore light microscope autoradiography has shown the movement of stigmatic proteins to the pollen but no penetration into the pollen was observed. The authors have hypothesized that the continuous cycling of glycoproteins is a kind of “second messenger”, which inhibits pollen probably by preventing the normal rehydration of its membrane. These glycoproteins are probably rapidly metabolized by pollen and therefore there is a necessity of their continuous synthesis, hence such an impact of cycloheximide.

The meaning of continuous protein synthesis was also noticed by Sarker et al. (1988). This study concerned pollen hydration and the phenomenon self-incompatibility, and was conducted by using metabolic inhibitors. The application of cycloheximide indicated that protein synthesis was necessary to regulate pollen hydration and self-incompatibility. Using tunicamycin, it was shown that the mechanism of self-incompatibility required some glycoprotein(s), which were not required for hydration control.

Extracellular protein coat of pollen was analyzed by Murphy (2006) in *A. thaliana*, *B. napus*, *B.oleracea* and *B. rapa* plants. The major component of the pollen coat is a specific group of proteins called pollenins, which comprise about 50–80% of all proteins. Murphy observed that there were many other proteins such as kinases, lipases, pectin esterase and caleosin.

Analyses of pollen grains were also conducted by Ruiter et al (1997). The main subjects of interest in this study were oleosins, which are glycine-rich proteins attached to the lipophilic part of pollen and absent in the remaining parts of the plant. Antibodies directed on the glycine-rich domain were prepared and used to detect proteins in extracts from sepal, carpel, pollen and stamen. This antibody detected several proteins in the pollen coat. The most abundant was 42 kDa protein. As the PCR analysis of genomic DNA and cDNA showed the existence of one and three different products respectively; the authors suggested that post-translational modifications were responsible for the several observed protein bands. In *Brassica napus*, oleosin-like proteins were always post-translationally modified. They had the N-terminal part removed. Protein bands after Western blot were observed in extracts from the pollen coat, pollen and whole stamen but not in the extracts from sepal, petal and carpel. In the fraction of the oil-body associated proteins isolated from pollen and stamen two oleosins were observed (42 and 34 kDa). After removal of the pollen coat by rinsing with cyclohexane no oleosins were detected, which confirms that oleosins are bound with the pollen coat. Additionally in the immature stamen 55 kDa protein was detected with the used antibody. The authors compared
the deduced protein sequences with those from *B. alboglabra* and *B. napus* and observed little differences in the glycine-rich domain. These differences result from deletions/insertions of repeating units. The glycine-rich domain is different from the other glycine rich motifs. It is randomly coiled, which allows the protein to bind large amounts of water. This feature could explain the fact that oleosins remain soluble after the boiling of the probe.

Other objects of interest in proteomic research were the proteins of the plant vascular system. Analysis of proteins (SDS-PAGE followed by MS/MS analysis) from the xylem sap (Buthz et al. 2004) of broccoli, rapeseed, pumpkin and cucumber showed the existence of 10–20 proteins with masses from 10 to 100 kDa. Overall protein concentration was 500-times lower than that of the proteins isolated from phloem. The most abundant proteins were identified as peroxidases, chitinases and serine proteases. They were observed in all of the analyzed species. The other identified proteins were: glycine-rich protein, lectin-like xylem sap protein 30, aspartyl protease family protein, subtilisin-like protein, lipid transfer protein, etc. Altogether 30 different proteins were identified. The protein patterns of Brassica plants (broccoli, oilseed rape) were very similar, although not identical. Obvious differences were observed on a family level (between Brassicaceae and Cucurbitaceae) as well as on the species level (between cucumber and pumpkin). The major components of xylem exudates in all species were various peroxidases. The authors found four peroxidases in broccoli, five in oilseed rape, two in pumpkin and three in cucumber. Besides the proteins observed in all species, the authors also identified proteins specific to individual plant species. The lectin-like xylem sap protein 10 was found only in cucumber; aspartyl protease was identified only in *B. napus* xylem and serine carboxypeptidase was identified only in broccoli exudates.

### 10.3.2 Proteomics of Stress Response

The second important area of proteomic research is the analysis of plant responses to biotic and abiotic stresses such as insects (Subramanian et al. 2005) and other pathogens (Peck et al. 2001) or wounding (Kim et al. 2003). Kim and coworkers (2003) analyzed the response of Chinese cabbage (*B. rapa ssp. pekinensis*) to wound stress and salicylic acid (SA). Proteins were extracted from the control and treated leaves of Chinese cabbage and analyzed by 2-D electrophoresis. The protein spots differentiating the two analyzed pools were excised from the gel and identified. The authors showed that SA and wounding induced pathogenesis-related protein 1a (PR1a) both on mRNA and protein level. PR1a was induced both by SA and by wound stress. Quantitative differences between the control and treated plants were observed. Many other proteins were enriched after application of SA. Nine of them were excised from the gel and identified by
MALDI-TOF MS after trypsin digestion. The identified proteins included for example: phospholipase D, homolog of resistance protein RPS2, arginase, serine/threonine kinase BNK1 and L-ascorbate peroxidase. Wounding induces the appearance of six proteins, namely: heat shock cognate protein 70, polygalacturonase and peroxidase P7, which are known to play a role in the defense system. Other induced proteins were: lysine-ketoglutarate reductase/saccharopine dehydrogenase, S-locus glycoprotein 12, and efflux carrier whose function in the defense system is unclear.

Besides drought stress, substantial losses in vegetable Brassicas crops are caused by pathogens. *Leptosphaeria maculans* is a devastating fungal pathogen of Brassica crops. It causes the so-called “backleg”. It is thus very important to understand the resistance mechanism of some plants to that pathogen. Subramanian et al. 2005 generated a cross between plants susceptible (*B. napus*) and resistant (*Brassica carinata*) to fungal pathogen (*Leptosphaeria maculans*) and analyzed the changes evoked in leaf proteomes. They used 2-D electrophoresis followed by tandem mass spectrometry. The results permitted identification of five proteins, unique to susceptible cultivar, (for example: fructose-bisphosphate aldolase, mutator-like trasposase, glutamine synthetase and triose-phosphate isomerase) and seven (carbonic anhydrase, thylakoid luminal 20 kDa protein, cyclin, B-type, sedoheptulose-bisphosphatase, nitrate reductase, hypothetical protein and superoxide dismutase) found only in a resistant genotype. They also observed 28 proteins differentially expressed in these genotypes in 48 hours after inoculation. The majority of the proteins identified in this study were antioxidant enzymes and enzymes involved in nitrogen metabolism, CO$_2$ fixation and photorespiration enzymes. Photosynthetic enzymes such as fructose biphosphate aldolase, triose phosphate isomerase, etc. were elevated in a resistant genotype after a pathogen attack. The proteins differentially regulated in these two genotypes were: malate dehydrogenase, putative cytochrome P450, rubisco large subunit, triose-phosphate, dehydroascorbate, hypothetical protein, unknown protein, fructose-bisphosphate aldolase, N-glyceraldehyde 2-phosphotransferase, sedoheptulose-bisphosphatase precursor, glutamine synthetase, ATP synthase (β-subunit), peptidyl-prolyl cis-transisomerase, topoisomerase IV subunit-β and peroxiredoxin type 2.

Liao et al (2008) studied the effects of di-N-butyl phthalate (DBP) —an endocrine disruptor on the growth of Chinese cabbage. Different concentrations (from 1 to 100 mg/L) of this compound were tested. In 2-DE analyses they noticed six protein spots differing in the control and the treated plants. They were: acyl-[acyl-carrier-protein] desaturase, root phototropism protein 3, ferredoxin-nitrite reductase, dihydroflavonol-4-reductase, aminoacyl-tRNA synthetase and ATP synthase (β-subunit). The first three showed increased expression, the rest decreased in response to the
stimulus investigated. The authors found that a closely related plant—Bok choy—responded differently.

Endocrine disruptors, and among them phthalate esters have been studied for years. They are widely used in industrial activities and exist commonly in the atmosphere, rain and soil. DBP is one of most popular phthalate ester and its half-life is 22 years in water solutions. Studies on a few species (Raphanus, Browalia, Capsicum and Brassica campestris) have shown that only DBP is absorbed and accumulated in plants, whereas plants do not take up the other phthalate esters. The biomass of the whole plant and chlorophyll content in leaves significantly decreased after DBP treatment. Changes in the protein content were analyzed by 2-D electrophoresis. Gels were stained with silver nitrate. Only six protein spots showed significant alterations. These proteins were identified by MALDI-TOF MS and peptide mass fingerprinting (PMF) analyses. Acyl-[acyl-carrier-protein] desaturase is involved in the biosynthesis of fatty acids (catalyzes the change of saturated to unsaturated acids), root phototropism protein is a signal transducer of phototropism response and ferredoxin nitrate reductase is involved in nitrate assimilation. Probably DPB could cause an increase in some physiological reactions. The level of the other enzymes (aminoacyl-tRNA synthetase, dihydroflavonol-4-reductase and ATP synthase) decreases in response to DBP, which can damage the cells and disturb the metabolism. The proteins, whose content decreases, are responsible for the biosynthesis of flavonoids, aminoacylation of tRNA and synthesis of ATP.

An interesting research concerning stress-specific proteome was conducted by Alvarez et al. (2009). They analyzed the root proteome of Brassica juncea in response to cadmium. Two methods were used: two dimensional difference gel electrophoresis (2D-DIGE) and isobaric tag for relative and absolute quantitation (iTRAQ). Each of the technologies used gave different sets of the proteins identified, however, these proteins belonged to similar functional categories. Proteins isolated from control and Cd-treated plants were labeled and separated on the same gel. The authors noticed an increase in the intensity of 26 and decrease in the intensity of 43 spots. In 69 spots with different intensity 102 proteins were identified. Up-regulated proteins were involved in sulfur assimilation, detoxification and redox homeostasis; while down-regulated ones were involved in protein synthesis and processing. By the other method 130 differently expressed proteins were identified, among them 16 were also identified by the 2D-DIGE method. The results have shown that on large scale analyses it is not only important to repeat the analysis (the authors made three replicates for each method) but the choice of a proper method or a combination of a few methods is also important. Similar observations were made by Schmidt et al. (2007). They obtained different protein pools depending on the isolation procedure.
Equally important as the protein level change, is the analysis of post-translational modifications in some proteins. Such changes have been studied, e.g., by Peck et al. (2001) who compared phosphorylation changes between the control and elicitor treated plants. The authors analyzed the perception of microbial signal molecules in *A. thaliana* by using radioactive orthophosphate and 2-D electrophoresis. They observed a protein, which was phosphorylated within a few minutes after treatment. This protein was described as Atphos 43. They also found that the phosphorylation of similar proteins of rice and tomato was also induced quickly after a pathogen attack. The fact that rice and potato respond similarly to elicitors and the close relationship of *A. thaliana* with Brassicaceae allows one to suppose that the response of these plants could be similar.

### 10.3.3 Other Analyses

Proteomic as well as metabolomic research has also been used in other analyses; for example to establish genetic relationships between Brassica varieties (Marquez et al. 2001; Taylor et al. 2002). Marquez and coworkers compared the proteins of etiolated seedlings (aerial part) of cabbage, mustard, rapes, radishes and *A. thaliana* to establish the genetic relationships between them. Studies of the etiolated material allows analysis of a greater number of low abundant proteins, as the probe contains a lesser amount of ribulose bisphosphate carboxylase/oxygenase, which accounts for nearly 50% of green tissue proteins. The authors decided to analyze only about 750 spots well defined and reproducible, located in the central part of the gel. As not every sequence mutation changes and also the size and/or isoelectric point of the protein, only a fraction of the allelic variations were detected. There were 149 proteins common to the Brassica species and 49 common to all species analyzed. After analyses and construction of the phylogenetic tree the authors concluded that there are two lineages: one between *B. oleracea* and *B. rapa* and the other between *B. nigra* and *Raphanus* sp. (Marquez et al. 2001).

Seed proteins as a support in the discrimination of cultivars were also analyzed by Mukhlesur et al (2004). They analyzed 32 cultivars of the *B. rapa* from Bangladesh, Japan and China. The authors observed about 32 protein bands in each sample and 10 of them were polymorphic. Differences between the cultivars were quantitative rather than qualitative.

Another analysis of genetic relationships on the protein level was performed on the seed storage proteins (Sadia et al. 2009). The authors analyzed the seeds of *B. napus*, *B. juncea*, *B. carinata* and *B. rapa*. The choice to analyze seed proteins was made as they are not sensitive to changes in environmental conditions. Because of a small amount of proteins the analysis could be conducted on SDS-PAGE. Thirty cultivars were analyzed...
and about 29-31 protein bands were detected in one sample. Thirteen of them differed in intensity between cultivars and only two were absent in some samples. For example in *B. rapa* there was one protein band about 33 kDa, in the *B. napus* cultivars the authors observed the protein of about 31 kDa, in *B. carinata* both bands were visible, whereas in *B. juncea* none of them were present.

Another research group (Albertin et al. 2005) studied proteome changes in autoploidical genome duplication by 2-D electrophoresis of proteins isolated from haploid, diploid and tetraploid *B. oleracea*. The authors analyzed the homozygous lines of *B. oleracea*, *B. oleracea* var. *italica* and *B. oleracea* var. *alboglabra*. The proteomes of two organs, leaf and stem, were analyzed. Altogether 1,700 protein spots were observed, among them 558 were common to all the samples analyzed. Comparison of leaf and stem proteomes showed quantitative (about 24%) rather than qualitative (about 18%) differences. Proteins showing differences were mostly of stem origin, they are stem-specific or stem overexpressed, but overall the stem spots had lower abundance than leaf ones. No distinct qualitative variations were observed between the forms at different ploidy levels. Differences between the two genotypes examined were lower than that between organs. Analyses showed that ploidy level is not a major cause of changes in protein expression. A comparison of diploid and tetraploid proteomes allowed the identification of four different spots, whereas the analysis of different plants (biological replication/expected significantly by chance) showed 11 differences. The authors concluded that modification at the ploidy level did not induce changes in green tissue proteomes.

Besides the analysis of protein content and their post-translational modification, the third important branch of proteomic research is the analysis of changes in enzyme activity. Unfortunately, there is a very small number of research reported in this area. One of the works concerning enzyme activities is that by Saroop and coworkers. They analyzed soluble and bound peroxidase activities in the course of *B. juncea* seed development (Saroop et al. 2002). Three hydrogen donors were used: ferulic acid, caffeic acid and chlorogenic acid. The peroxidase activity rises after 20 days after anthesis. A reverse correlation was observed between peroxidase activity and water content. It suggests that this enzyme is involved in the termination of cell elongation. Changes in soluble peroxidase activity increased sharply between 20–30 days after anthesis, later the activity decreased. Levels of activity differed depending on the substrate. For ferulic acid minimal activity was observed and for caffeic and chlorogenic acid very similar results were obtained. The bound peroxidase activity was lower than that of the soluble one when the hydrogen donor was caffeic or chlorogenic acid. When ferulic acid was used, the activities were the same. The authors did not analyze how many isoenzymes were responsible for the activity of each
fraction. Another article devoted to the catalytic properties of enzymes is the work of Tayefi-Nasrabadi (2008) concerning catalases from *B. oleracea* var. *gongylodes* (kohlrabi). Catalases are an important class of antioxidant enzymes participating in practically every stress defense. The author extracted protein from kohlrabi bulbs and performed kinetic analyses. He identified three catalase isoenzymes. These isoenzymes had a different optimal pH (4.5, 6.5, and 10) and the most efficient was the isoenzyme active at pH 6.5. The isoenzyme active at pH 4.5 was more sensitive to azide and more resistant to cyanide than the other ones. It also showed competitive inhibition. Three isolated isoforms also differed in the heat inactivation. The isoform active at pH 6.5 was inactive above 50°C; and the isoform active at pH 10 was inactive above 70°C. The differences in the effects of pH, temperature, substrate and inhibitors on the catalases activity suggest different roles of these isoenzymes.

Among the papers devoted to a few proteins, an interesting one is that by Köksal and Gülçin (2008). The authors purified and characterized peroxidase from cauliflower buds. Peroxidases are a class of enzymes, which catalyze the oxidation of many substrates in the presence of hydrogen peroxide. They are present in many organisms and are the key enzymes in control of plant growth and development. The enzyme described was isolated from crude protein extract by ammonium sulfate fractionation and subsequent ion exchange chromatography (CM-Sephadex A-50). The molecular mass of the isolated enzyme was determined by gel filtration chromatography as 44kDa. Electrophoresis in native polyacrylamide gel showed only one band—so there was only one isoenzyme. The optimum pH for this isoenzyme was 7.5 and the optimal temperature varied depending on the substrate used; it was 25°C for pyrogallol, 30°C for guaiacol and 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid (ABTS) substrates, 45°C for 4-methyl catechol, 50°C for catechol. Kinetic studies showed that this peroxidase had the highest affinity to pyrogallol. The authors suggested that this enzyme can be used for the synthesis of some aromatic compounds, removal of phenolic peroxides from water and industrial wastes. The literature devoted to proteomic research is scarce and concerns mainly other issues (different protein pools and stresses), so it is difficult to find any common denominator such as the same protein or type of stress. Quite a lot of articles are concerned with the differentiation of species and cultivars on the basis of protein profile. Unfortunately, the authors did not identify most of these proteins and only compared the profiles. From other analyses we can find out that most of the identified proteins belonged to the functional classes such as defense proteins (especially antioxidant enzymes) and metabolic enzymes. Surprisingly almost no proteins connected with signalization were identified. Only Alvarez et al. (2009) clearly stated that the level of some proteins from the signaling pathways differed between
the control and the cadmium treated plants. Further analyses are needed to recognize the whole proteome of the vegetable Brassicas. Undoubtedly, the proteomic analyses of *A. thaliana* and *B. napus*, as well as recognition of the *B. campestris* genome will be a great help.

### 10.4 Current Status of Vegetable Brassica Metabolomics

Metabolomics has many advantages compared to transcriptomics and proteomics (Taylor et al. 2002). Firstly, a single change in the transcript or a protein can cause alterations in many metabolites or a substantial change of some metabolites. Such amplification of changes facilitates their detection. Secondly, metabolomic research does not need a full sequence of genome, as is needed in transcriptomics and proteomics research. Thirdly, metabolomics is more generic as genes and proteins differ between species and most metabolites are the same in all organisms. Four different approaches are used in metabolome analysis: target analysis, metabolite profiling, metabolomics and metabolic fingerprinting (Fiehn et al. 2001). Target analysis is the study of the substrate or direct product of one specific protein. Metabolite profiling (also called metabolic profiling) is the analysis of a pre-defined group of metabolites, for example members of particular metabolic pathways. Metabolomics concerns the identification and (sometimes) quantification of all metabolites from one cell, tissue, organ, etc. Metabolic fingerprinting (sometimes called metabonomics) is a comparison of some metabolite/metabolites between genotypes/lines/probes. Each area of research requires different and strictly controlled methods of isolation and analysis of metabolites. Quite frequently proteomic research is accompanied with some metabolite analysis. Extracellular proteins from pollen coat were analyzed by Murphy (2006) in *A. thaliana*, *B. napus*, *B. oleracea* and *B. rapa*. Besides proteins, the second group of major components of pollen coat is lipids enriched in acylated compounds. They have a unusually high degree of saturation and different acyls. In coat protein there are mainly, campesterol, and campestdienol, while in lipids we usually observe acyls such as myristate, palmitate, and stearate. This lipidic layer protects male gamethophytes from dehydration. Metabolomic research is more diverse than the proteomic one. However, a large number of these analyses concern the response to various stresses. In relation to Brassicaceae, a large part of research is also connected with analysis of glucosinolate content.

#### 10.4.1 Glucosinolates and Related Metabolic Compounds

The majority of metabolic analyses in Brassica vegetables are the metabolite profiling type, and the main subjects of interest are glucosinolates (GLS). All glucosinolates characterized (β-thioglucoside −N-hydroxysulfates)
share a similar structure: β-D-glucopyranose with the thio group and chain derivative from amino acid (methionine, alanine, leucine, phenylalanine or tryptophane) (Fig. 10-3). More than 120 chemically distinct glucosinolates have been identified in plants. The majority of GLS have been isolated from crucifer family members, and about 15 other plant families are known to contain glucosinolates, but not all are edible (Akaniaceae, Bataceae, Bretschneideraceae, Capparaceae, Caricaceae, Drypetes Euphorbiaceae, Gyrostemonaceae, Limnanthaceae, Moringaceae, Pentadiplantdraceae, Resedaceae, Salvadoraceae, Tovariaceae, Tropaeolaceae) (Rodman 1991a, b). For more detailed characterization of glucosinolates also see Chapter 1.

Figure 10-3 Structure of glucosinolate (http://en.wikipedia.org/wiki/Glucosinolates).

In literature we can find analyses of GLS content in different cultivars or developmental stages (Bayoumi et al. 2007; You et al. 2008), works on the impact of storage and preparation of vegetables on GLS digestibility (Rungampestry et al. 2007) and changes of these compounds in various environmental conditions (Liang et al. 2006; Widarto et al. 2006; Wurst et al. 2006; Jahangir et al. 2008).

Research conducted by Bayoumi et al (2007) concerned the metabolomic analysis of *B. rapa* (various cultivars). The most diverse cultivar metabolites were amino acids, carbohydrates, adenine, indole acetic acid, phenylpropanoids, flavonoids, and glucosinolates. The authors analyzed four cultivars of *B. rapa* (Raapstelen, Herfstraap, Witte Mei, and Oleifera) at two development stages (4 and 6 weeks old). The level of phenylpropanoids and glucosinolates was higher in older plants. As these compounds are involved in defense processes, it is suggested that there is need to increase the defense potential. In younger plants, the levels of amino acids and organic acids are higher. The lower level of these molecules in older plants can be caused by the fact that they are necessary for glucosinolates and phenylpropanoids synthesis.

Derivatives of GLS, generated by hydrolysis with myrosinase (Fig. 10-4), such as thiocyanates, isothiocyanates, nitriles, epithionitriles are involved in plant defense against various pathogens.

Isothiocyanates (ITC), such as sulforaphane (SNF) or its analogue erucin (Fig. 10-5) also have pharmaceutical meaning as anticancer agents.
On the other hand, some compounds resulting from the hydrolysis of glucosinolates can have harmful effects on human health. For example goitrin (Fig. 10-6) reduces the production of thyroid hormones and in some conditions it can form N-nitrosooxazolidone, a mutagen. For this reason You et al (2008) searching for the best source of sulforaphane (as a human diet complementation) also analyzed the goitrin content.

![Figure 10-4](http://lpi.oregonstate.edu/ss06/vegetables.html) Hydrolysis of glucosinolates to its isothiocyanate by myrosinase.

![Figure 10-5](http://en.wikipedia.org/wiki/Goitrin) Structure of goitrin.

![Figure 10-6](http://en.wikipedia.org/wiki/Goitrin) Structure of goitrin.

The HPLC analysis of seeds of 43 different *B. oleracea* varieties (You et al. 2008) showed that erucin was present mainly in cabbage (*B. oleracea* var. *capitata*) and kohlrabi (*B. oleracea* var. *gongylodes*) and sulfophorance in broccoli (*B. oleracea* var. *italica*). The content of isothiocyanates also depended on the climate and other environmental factors. The authors analyzed 43 cultivars belonging to 10 *B. oleracea* species. Among them only 21 cultivars contained detectable levels of sulfophorance and erucin. In broccoli, cabbage and kohlrabi there were significant differences in the isothiocyanates content between cultivars. When the seeds started to sprout the isothiocyanates level decreased sharply. Seeds of the *Brassica* species contain 10 times more isothiocyanates than the edible parts but they are toxic to humans and
animals; so they can only be a source for the extraction of metabolites. The authors analyzed the level of sulforaphane, erucin and goitrin in seeds. They suggested using Brassica seeds as a diet complementary or as a source for the extraction of isothiocyanates. On the basis of the results obtained the most suitable source of food supplements were brocoli seeds. They contained a high level of sulforaphane and some cultivars (such as Green Dragon II) also had low levels of goitrin.

For Brassica (and other vegetables) some results are available on the impact of storage and cooking on metabolite composition. One of the works concerning glucosinolate level is by Rungampestry et al. (2007). Besides the environmental changes, the glucosinolate level is also affected by storage, processing and ingestion of vegetables. It is known that glucosinolates are hydrolyzed by plant myrosinases after tissue damage. In normal cells glucosinolates and myrosinase are spatially separated (placed in different compartments); after damage (for example by insects) myrosinase is released and glucosinolates are hydrolyzed to toxic compounds (such as isothiocyanates). Cooking can also change the glucosinolates level by impact of myrosinase activity. Besides the glucosinolate content, their anticancer activity also depends on the mode in which they are taken up by the organism. Eating raw Brassica vegetables causes the release of glucosinolates in the upper gastrointestinal tract; this releases the myrosinase present in plant tissue. Glucosinolates from cooked vegetables are released in the colon with the involvement of the colon microflora. Changes in the level of glucosinolates/isothiocyanates can also be caused by the introduction of transgenes. Zang et al. (2007) introduced three cDNAs coding cytochrome 450 family enzymes from Arabidopsis thaliana to Chinese cabbage. These are CYP79B2 and B3 from the biosynthetic pathway of indole-3-acetic acid (IAA) and indole-glucosinolate (IG) and CYP83B1 unique to IG biotynthesis. They analyzed the glucosinolate content of three independent lines transformed with each double and triple construct. They found that the lines containing CYP79B3 and CYP83B1 have the same level of IG as controls. Overexpression of CYP79B3 and CYP79B2 causes IAA overproduction that prevents the regeneration of callus. There were no indole glucosinolates in these lines which confirms the role of these enzymes in IG biosynthesis. The lines containing CYP79B3, CYP79B2 and CYP83B1 regenerated normally, and the obtained plants had elevated levels of glucobrassicin, 4-hydroxy glucobrassicin, and 4-methoxy glucobrassicin. Overexpression of CYP79B3 alone did not alter the IG level. HPLC analyses showed that the main differences in IG accumulation in transgenic plants concern glucobrassicin, while neoglucobrassicin (56% of all IG) was not affected. The authors conclude that it is possible to enhance the IG biosynthesis without affecting plant development by the simultaneous overexpression of three enzymes from this pathway. As mentioned above, glucosinolates are the
main subjects of interest in Brassica metabolomics. Changes of their level are also analyzed in different stresses and developmental stages, which is described later in the text.

10.4.2 Metabolite Changes in Stress Conditions

Phospholipid catabolism in cauliflower treated with $\gamma$-irradiation has been studied by (Voisine et al. 1993). The authors observed the acceleration of membrane deterioration caused by this treatment, so they investigated the activity of lipolitic enzymes. The breakdown products obtained suggested that $\gamma$-irradiation changes the activity of the membrane bound phospholipase D, phosphatidic acid phosphatase and lipolitic acyl hydrolase. The enhancement of membrane deterioration is induced mainly by free radicals.

Recently, a report on plant interactions with below-ground biota was published (Wurst et al. 2006). The authors investigated changes in metabolite composition in $B. oleracea$. They analyzed the combined impact of nematode ($Meloidogyne incognita$) and earthworm ($Octolasion tyrtaeum$) and reported an increase in the N uptake and shoot biomass. As for glucosilonates, glucoiberin level was decreased in the presence of earthworms and the content of glucoraphanin was affected in the presence of both organisms, but the decrease in this content was more pronounced in the presence of each organism separately than when these two organisms were both present in the soil. The contents of the other glucosilonates (sinigrin, progoitrin, glucobrassicin, 4OH-glucobrassicin, MeOH-glucobrassicin, neoglucobrassicin) were not affected. Changes in the level of glucosilonates (especially glucoiberin) may affect the oviposition behavior of insects. The results obtained demonstrated that the presence of earthworms and/or nematodes in the soil can increase N availability and hence reduce the glucosinate levels. On the other hand, changed glucosinate levels can affect the above ground herbivore performance. The analyses performed by Liang et al. (2006) have shown that in $B. rapa$ leaves treated with methyl jasmonate (MJ) changes in the contents of several metabolites were observed. They noticed a decrease in the content of sugar and amino acids, while the level of hydroxycinnamates and glucosilonates increased substantially. Treatment with MJ causes IAA and IG accumulation. It is known that MJ plays a role in a variety of biological processes including defense responses. The results obtained suggest that the components of key importance in the defense system of $B. rapa$ leaves can be the metabolites such as hydroxycinnamates, indole derivatives and glucosilonates.

Industrial development has been increasing the contamination of soil with different metal ions. It is known that Brassica plants accumulate metals in large amounts; they are even used for the phytoremediation of
contaminated soils (Jahangir et al. 2008). B. rapa plants were treated with increasing concentrations (50–500 mM) of three metal ions (C, Fe and Mn). After treatment metabolome changes were analyzed by NMR followed by chemometric methods. The authors observed changes in the contents of key metabolites, glucosinolates and hydroxycinnamic acids. Besides the metabolite changes after metal treatment, they also observed substantial differences in the components of roots and shoots. In roots they found large amounts of sugar, glucosinolates and some free amino acids, while in leaves there were larger amounts of phenylpropanoids. Regarding these differences metal impact was analyzed separately in roots and leaves. After treatment, the sugar levels decreased, which suggests that metal ions affect photosynthesis. A great number of changes were similar for all the metals considered, however the authors observed some differences. For example, when 100 mM Fe is used, substantial changes in the content of alanine, threonine, valine and glutamic acid differences are observed; while the addition of 100 mM Cu caused differences in serine concentration. Copper and iron have greater effect than manganese. As a result of the treatment with all the metals considered, the contents of amino acids and phenylpropanoids increased, which could be a detoxification response. It is known that amino acids and phenolics have metal-chelating properties. The changes mentioned above were observed for low and moderate metal concentrations. High concentrations (~ 500 mM) cause a decrease in the content of all the metabolites identified, such concentrations are toxic to plants. Plants tolerate metal to some extent, and with increasing metal concentrations they increase the level of some metabolites, such as amino acids and phenolics. However, when the metal concentration exceeds some critical point, the metal ions interfere with biosynthesis pathways causing a decrease in the contents of all metabolites and reduction of growth. The response of plants depends not only on the metal ion concentration but also on its type (Jahangir et al. 2008). A group of metabolites frequently found to change in response to various stresses are lipids. Nouairi et al (2006) examined the impact of cadmium on the membrane lipid content in B. napus and B. juncea plants. Lipid changes depended on cadmium concentration, but most of the differences were observed between species. Total lipid content decreased significantly in B. napus, while B. juncea showed a totally different reaction. Also the level of malondialdehyde (as an indicator of the peroxidation level of membranes) and that of nonproteinaceous thiol (as a chelation agent) were measured. The impact of cadmium on lipid biosynthesis was analyzed by using of radiolabeled acetate. The authors observed lowered levels of lipids and their biosynthesis in B. napus leaves and increased levels of lipid peroxidation (6 times more of malondialdehyde in comparison to that in the control plants). There were mainly changes in plastids membranes (decreased level of glycolipids) with no or little changes
in the extrachloroplastic lipids (slightly increased levels of phospholipids). In *B. juncea* lipid biosynthesis was stimulated by cadmium and there are no differences in the malondialdehyde content between the control and treated plants. Such stimulation of lipid biosynthesis can be explained by the fact that membranes are essential for the formation of vesicles involved in the compartmentalization of heavy metals. Changes in thiol concentration, which is important in chelating heavy metals, also support this fact. In oilseed rape there is no change in the thiol level, and in *B. juncea* the authors observed a significant increase (up to 20 times of the control level in leaves and up to 70 times in roots) in the content of thiol after cadmium treatment. This result explained the resistance of *B. juncea* to heavy metals and suggests that this plant could be a good phytoremediator.

Changes in *B. rapa* leaves metabolome after the attack of Lepidoptera (*Plutella xylostella* L. and *Spodoptera exigua* Hubner) were investigated by NMR (Widarto et al. 2006). The response induced by the herbivore attack is marked by production of a number of secondary metabolites. The metabolic changes can be specific depending on the attacking species or common to all treatments. After attack by insects, the plants at first initiate wound-response within a few minutes and then differences in signal transduction stimulate a specific response to a specific pathogen (Widarto et al. 2006). It was demonstrated earlier that herbivore attack causes changes in the glucosinolate level, some alkaloids and a wide variety of volatile organic compounds in plants. The study by Widarto et al. (2006) demonstrated that the metabolic changes caused by the 2nd and 4th instar larvae were different. In the leaves infested by the 2nd instars the increased levels of glucose, ferulic acid, and gluconapin were noted, whereas the leaves attacked by the 4th instars had increased levels of alanine and sinapoyl malate. In both tested species metabolome changes caused by young larvae were greater than those caused by older ones. It could be caused by the fact that younger larvae are not so mobile and heavily damage leaves in isolated places.

### 10.4.3 Other Analyses

As seen in the literature on proteins, the papers on metabolomics present great diversity of subjects. A metabolomics-based approach to plant genotype discrimination was proposed (Taylor et al. 2002). Two genotypes of *A. thaliana* (Col0 and C24) and their progeny that differ only in mitochondrial and chloroplast genotypes were analyzed. Two crosses were made differing in maternal inheritance. Only the polar phase isolated from leaves was analyzed. The authors used gas chromatography coupled with mass spectrometry (GC/MS) and identified 433 metabolites in the samples. Using the descriptive statistics they showed that the level of the key primary metabolites varies more than others. The information obtained
was sufficient to discriminate between parental lines and each of the F\textsubscript{1} progeny. Both, metabolomic (Taylor et al. 2002) and proteomic analyses (Marquez et al. 2001; Mukhlesur et al. 2004; Sadia et al. 2009) are often used to characterize cultivars/germplasm collections. Metabolic research also includes the analysis of mineral content. Analyses conducted by Broadley and coworkers showed that mineral concentration differed significantly between varieties. The authors analyzed the content of Ca and Mg in 376 varieties of *B. oleracea* and its wild relatives. It was demonstrated that the shoot Ca and Mg content was heritable (up to 40%) and mapped quantitative trait loci (QTL) on chromosomes C2, C6, C7 and C8. The results proved that the shoot Ca and Mg content varies between analyzed lines and did not correlate with plant biomass (Broadley et al. 2008). The content of Ca varied from 1.7 to 3.3\% and that of Mg from 0.35 to 0.8\%. The highest level of these ions was observed in *B. oleracea* var. *capitata* and lowest in *B. oleracea* var. *gongylodes*. Although aflatoxins (AF) do not belong to the metabolites produced by plants, they can be present in their edible part, so it is important to analyze the conditions in which the aflatoxin content increases. Aflatoxins contamination of food is one of the most serious problems concerning human health; they can also cause liver cancer. One of the major factors increasing AF content in plants is drought stress (Guo et al. 2008). High temperature also increases AF content. Aflatoxins are produced mainly by *Aspergillus flavus*. Comparison of the susceptible and resistant cultivars has shown that susceptible lines induce the same proteins as resistant ones but at a lower level or much later after the pathogen attack. Differences affected pathogen induced proteins as well as constitutively expressed ones. Tobacco plants overexpressing glyoxalase I from *B. juncea* were more resistant to *Asperigillus*. This enzyme probably acts by the removal of methylglyoxal, which is the AF inducing substrate.

**10.4.4 Metabolite Changes in the Course of Development**

Changes in antioxidant compounds during the seed sprouting of *B. oleracea* var. *costata* were analyzed (Sousa et al. 2007). Phenolic compounds such as the esters of sinapic acid and organic acids (conitic, citric, pyruvic, malic, shikimic, and fumaric) were identified. The authors observed a decrease in phenolic compound levels but no qualitative changes. Total organic acid content increased but some of the compounds (aconitic, citric and shikimic acid) occurred at a decreased level. Chromatographic analysis with diode array detector (HPLC/DAD) revealed the presence of 12 phenolic compounds in germinating sprouts. They included five sinapoyl glucosides, four disinapoyl glucosides, one trisinapoyl glucoside, sinapoyl choline, and one kaempferol glucoside acylated with sinapic acid. The total phenolic content showed significant quantitative changes. Its level decreased from
11.1 g/kg in 2-day-old sprouts to 1.6 mg/kg after 10 days. This decrease could be explained by using phenolics in cell wall biosynthesis and as antioxidants. Contrary to phenolic compounds, total organic acids content increased from 45.7 g/kg on day 2 to 66.9 g/kg on day 12. The main organic acids found at the time of germination are malic and citric acids, although the content of the latter decreased slightly from day 2 to 12. Despite the total organic acids increase, there are changes in the quantity of individual compounds at the time of sprouting, most of them in the first 4 days. As glucosinolates (including derivatives of their hydrolysis) have great impact on human health (especially as anti-cancer agents) the majority of the metabolic analyses based their estimation in different *Brassica* species and different environmental conditions.

10.5 Applications of Proteomics and Metabolomics Data (Clinical Aspects: Anti-Carcenogenic Effects—Human Proteome Research)

Cruciferous vegetables are commonly used in the human diet and contain many bioactive compounds, e.g. flavonoids, vitamin C, K, minerals such a selenium and glucosinolates (Moreno et al. 2006). The nutritional properties of Brassica crops and their use in human diet are described in Chapter 1. It has been shown that all of these biomolecules present in cruciferous vegetables may be protective natural compounds and should help prevent development of many cancers (Li et al. 2008; Boivin et al. 2009). The most intensely studied cruciferous biomolecules associated with cancer prevention have been glucosinolates. GLS are not bioactive, they must be first enzymatically hydrolyzed (by myrosinase) to isothiocyanates (Fenwick et al. 1983) (described previously in 10.3, Fig. 10-7). There are two most important and well-studied isothiocyanate members. The first one is sulforaphane, which was isolated and characterized, in the early 1990s mostly in broccoli, Brussels sprouts, cauliflower and kale (Zhang et al. 1992) and the second one is indole-3-carbinol (I3C) detected mainly in broccoli and cabbage (Moreno et al. 2006). Many studies conducted on animals (rodents) or human cells have shown that there are several mechanisms in which sulforaphane and isothiocyanates could act as cancer protective molecules. These include: 1) induction of Phase 2 detoxification enzymes, and 2) inhibition of carcinogen-activating Phase 1 enzymes, 3) induction of apoptosis, and 4) regulation of cell progression via cell cycle arrest, and 5) modulation of hormone metabolism, 6) also anti-inflammation. For a more detailed discussion of target genes and mechanisms of glucosinolates action the reader is directed to several extensive review articles (Bheemreddy and Jeffery 2006; Myzak and Dashwood 2006; Fimognari and Hreila. 2007; Clarke et al. 2008). Here we focus only on a few examples of glucosinolates’ cancer protective mechanisms concerning mainly human cancer.
Jackson and Singletary (2004) have shown that sulforaphane can stop the growth of human breast cancer cells and identified the mechanisms of action of this compound. They discovered that SNF can stop cancer cells development by interfering with microtubule assembly, which are necessary for cell division and consequently inhibit the growth of tumor cells. These anti-proliferative effects of sulforaphane have been compared with cancer therapeutic drugs in the study reported by group led by Merlin Morris. The study has shown that higher concentration of SNF could stop cancer cell proliferation comparable to the effects of drugs, e.g., daunomycin or vinblastine (Tseng et al. 2004). Singh and coworkers have shown that sulforaphane can also inhibit the proliferation of the human PC-3 prostate cancer via inducing caspase-mediated apoptosis (Singh et al. 2004). Also Shan and colleagues have discovered that sulforaphane can induce apoptosis in the human bladder cancer cell (Shan et al. 2006). Furthermore, this compound could also inhibit the growth of several of the *Helicobacter pylori* strains, bacteria associated with gastric cancer development (Surh et al. 2001; Fahey et al. 2002).

<table>
<thead>
<tr>
<th>Category</th>
<th>Gene</th>
<th>Change in activity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phase I enzyme</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP1A1</td>
<td>Decrease</td>
<td></td>
</tr>
<tr>
<td>CYP2B1/2</td>
<td>Decrease</td>
<td></td>
</tr>
<tr>
<td>CYP3A4</td>
<td>Decrease</td>
<td></td>
</tr>
<tr>
<td>CYP4A10</td>
<td>Increase</td>
<td></td>
</tr>
<tr>
<td>CYP4A14</td>
<td>Increase</td>
<td></td>
</tr>
<tr>
<td>CYP39A1</td>
<td>Increase</td>
<td></td>
</tr>
<tr>
<td><strong>Phase II enzyme</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NQO-1</td>
<td>Increase</td>
<td></td>
</tr>
<tr>
<td>GST</td>
<td>Increase</td>
<td></td>
</tr>
<tr>
<td>γ-GCS</td>
<td>Increase</td>
<td></td>
</tr>
<tr>
<td>UGT</td>
<td>Increase</td>
<td></td>
</tr>
<tr>
<td><strong>Anti-oxidants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thioredoxin reductase 1 and 3</td>
<td>Increase</td>
<td></td>
</tr>
<tr>
<td>Glutathione peroxidase 3</td>
<td>Increase</td>
<td></td>
</tr>
<tr>
<td><strong>Proteasome subunits</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha 1 and 3, beta 5, and 26S</td>
<td>Increase</td>
<td></td>
</tr>
<tr>
<td><strong>Cell cycle and cell growth</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclins D1, E2 and T2</td>
<td>Increase</td>
<td></td>
</tr>
<tr>
<td><strong>Stress response</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heat shock proteins 1, 1 alpha,</td>
<td>Increase</td>
<td></td>
</tr>
<tr>
<td>8 and 105</td>
<td>Increase</td>
<td></td>
</tr>
<tr>
<td><strong>Transcription factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CREB binding protein</td>
<td>Increase</td>
<td></td>
</tr>
<tr>
<td>Aryl hydrocarbon receptor nuclear translocator-like</td>
<td>Increase</td>
<td></td>
</tr>
</tbody>
</table>

Figure 10-7 A. Schematic representation of glucosinolates hydrolysis. Reprinted from Bheemreddy and Jeffery (2006). B. Molecular targets of sulforaphane. Redrawn according to Fimognari et al. (2007). C. Effects of sulforaphane on “blocking” mechanisms. Redrawn according to Clarke et al. (2008).
Another representative of thiocyanates, indole-3-carbinol (I3C) in the human body is mostly converted into 3,3’-diindolylmethane (DIM), which is the bioactive dimer of I3C (Aubom et al. 2003). Several reports have shown that this compound also acts in multiple ways against the development of cancer. Firstly I3C could negatively regulate estrogen hormone metabolism (Aubom et al. 2003). Estrogen hormones promote cancer cell growth and DIM may suppress it and consequently prevent the development of estrogen-enhanced cancer such as breast (Bradlow et al. 1991; Meng et al. 2000), cervical (Jin et al. 1999) or endometrial (Kojima et al. 1994) cancers. DIM can also inhibit the growth of prostate cancer cells via suppression of the proteins that speed up the growth of cancer cells (Smith et al. 2008). It was also shown that indole-3-carbinol inhibits the action of most prominent carcinogens derived from cigarettes called nitrosamines (e.g., NNK, nicotine-derived nitrosamine ketone, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butane). I3C may reduce the incidence of DNA damage caused by nitrosamines through changes or inhibition of its metabolism (Hecht et al. 1996; Taioli et al. 1997). Apart from the glucosinolates the most important biomolecules present on a high level in cruciferous vegetables are also carotenoids (lycopene), polyphenols, fibers and vitamins, especially vitamin C (Moreno et al. 2006), which all may act in cancer prevention as excellent antioxidants and additional stimulators of the immune system (Ferguson and Harris 1996; Yang et al. 1997; Sikora et al. 2008; Boivin et al. 2009).

Because of the rapid progress in high-throughput techniques over the past 20 years it was possible to demonstrate that the consumption of fruits and vegetables, especially cruciferous vegetables can reduce incidences of many types of cancer (Talalay and Zhang 1996; Hecht 1999). This progress facilitates better understanding of the role of various compounds in human health and helps to determine their proper amounts in the human diet (Talalay and Zhang 1996; Mithen et al. 2000).

Acknowledgements

This study was supported by the Polish Committee for Scientific Research Grants PZ-19/2006/2/JS and PBZ-MNiSW-2/3/2006/19/IGR/4.

References


Role of Bioinformatics as a Tool for Vegetable *Brassica* Species

*Chris Duran,† Zoran Boskovic,† Jacqueline Batley,‖ and David Edwards†.*

**ABSTRACT**

With the rapidly increasing volumes and complexity of genetic and genomic data available for *Brassica* crops, there is an increasing need for bioinformatics tools to analyze and organize this data. This chapter reviews a selection of tools and databases available for *Brassica* bioinformatics, describing their purpose and utility for *Brassica* researchers. The majority of *Brassica* bioinformatics tools have been developed for the discovery and analysis of molecular genetic markers. These include autoSNPdb, SSRPrimer CMAP3D and Brassica.info. With the growth of gene and genome sequence data, as well as gene expression and phenotypic information, there will be an increasing requirement to integrate and merge this information using current and future bioinformatics applications.

**Keywords:** autoSNPdb, SSRPrimer, CMap, EnsEMBL, Brassica.info

### 11.1 Introduction

With the continued development of molecular techniques, the quantity and variety of data obtained by researchers is growing at an exponential rate. This increase in data production makes data analysis and integration a major challenge and increasingly sophisticated methods are required to draw meaningful conclusions from the analysis. It is becoming extremely
Role of Bioinformatics as a Tool for Vegetable Brassica Species

It is difficult for humans to interpret the experimental results without the use of dedicated bioinformatics tools, which enable the management and analysis of large quantities of information.

The demand for bioinformatics tools has increased with increasing data and is leading to a virtual revolution in science, with an increasing focus on data analysis rather than data production. With the escalating complexity of bioinformatics analysis, there is the potential to use the wrong tool, or to misinterpret data. It is, therefore, important that researchers are aware of the capability and limitations of the tools being applied as well as limitations related to data quality. In this chapter, we describe a range of bioinformatics tools and databases available to Brassica researchers that may be applied to assist molecular genetic analysis in these species leading to improved Brassica varieties.

Bioinformatics systems fall into several categories, including central data repositories such as the Protein Data Bank (PDB, http://www.rcsb.org/pdb), which contains information about experimentally-determined protein structures, nucleic acids and complex assemblies; the National Centre for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/), a repository for DNA sequence information; and UniProt (http://www.uniprot.org/), a protein sequence and functional information resource. In addition to central repositories, there are several more specialized databases, such as AutoSNPdb (Duran et al. 2008) and CMap (Lim et al. 2007), as well as general data analysis tools, such as the molecular marker discovery tools, SNPServer (Savage et al. 2005), SSRPrimer (Robinson et al. 2004) and SSR Taxonomy Tree (Jewell et al. 2006); the sequence clustering tool CLUSTAL-W (Larkin et al. 2007); and the sequence comparison tool BLAST (Basic Local Alignment Search Tool; Altschul et al. 1990). There are also general Brassica databases such as brassica.info (http://www.brassica.info/) and the Brassica Genome gateway (http://brassica.bbsrc.ac.uk/); and comparative genomics tools such as: Brassica—Arabidopsis Comparative Genome Viewer (http://brassica.agr.gc.ca/navigation/viewer_e.shtml), the AtEnsembl viewer (http://atensembl.arabidopsis.info) and the ATIDB database resource (http://atidb.org).

This chapter reviews the available Brassica tools, detailing their purpose and utility for scientists and breeders.

11.2 Genetic Resources

Genetic studies examine genetic variation in relation to heritable traits. The majority of variation between individuals is found at the nucleotide level, and molecular genetic markers are powerful tools for the association of nucleotide variation with heritable traits. Two forms of sequence based markers, simple sequence repeats (SSRs), also known as microsatellites, and single nucleotide polymorphisms (SNPs) have become the major markers...
currently applied in modern genetic analysis. These are supplemented with anonymous marker systems such as amplified fragment length polymorphisms (AFLPs; Vos et al. 1995) and diversity array technology (DArT; Wenzel et al. 2004). Molecular genetic markers are used for genetic mapping and the identification of genes underlying heritable traits, which makes them excellent tools for crop improvement by marker-assisted selection.

11.2.1 Marker Discovery

The application of molecular genetic markers for crop improvement has traditionally been limited by the huge costs related to marker discovery and association with valuable agronomic traits. Early marker discovery methods were laboratory based, expensive and time consuming. However, with the increased application of high-throughput sequencing technology, there is a vast amount of DNA sequence data available in the public domain, allowing bioinformatics developers to apply computational methods for the identification and annotation of large numbers of markers rapidly and at low cost. Having access to large numbers of markers, such as SNPs and SSRs, at a high density in the genome provides a powerful resource for molecular geneticists. SNPs and SSRs are routinely used in animal and crop breeding programs (Gupta et al. 2001), for genetic diversity analysis, cultivar/breed identification, phylogenetic analysis, characterization of genetic resources and association of genetic loci with valuable traits (Rafalski 2002). Several tools have been developed for the discovery of SNPs and SSRs from DNA sequence data and a selection of these is listed below.

11.2.1.1 AutoSNPdb

AutoSNPdb combines SNP discovery software and sequence annotation with a relational database (Duran et al. 2008). The software implements the autoSNP SNP discovery pipeline (Barker et al. 2003; Batley et al. 2003) and a custom sequence annotation pipeline within a relational database to enable mining for SNP and indel polymorphisms associated with genes of known function or comparative physical map position. A non-redundant set of *Brassica* gene sequences were mined for SNPs, annotated by comparison with GenBank and UniRef90, and compared with the complete *Arabidopsis* genome sequence. A web interface enables searching and visualization of the data, including the display of sequence alignments and SNPs (Fig. 11-1). The tool allows users to identify SNPs between *Brassica* cultivars or within genes of predicted function. The database hosts details of 203,036 predicted *Brassica* SNPs in 169,721 unigene sequences and can be accessed at: http://acpgf.imb.uq.edu.au/autosnpdb.php.
11.2.1.2 SSRPrimer

SSRPrimer (Robinson et al. 2004; Jewell et al. 2006) provides an automated process to identify and design PCR primers for SSR loci by combining the SSR discovery tool Sputnik and the PCR primer design software Primer3. The scripts take multiple sequences in FASTA format as input and produce lists of SSRs and associated PCR primers in tabular format. This web-based tool is also available as a stand alone version for very large datasets. SSRPrimer has been successfully applied to a wide range of species including shrimp (Perez et al. 2005), citrus (Chen et al. 2006), mint (Lindqvist et al. 2006), strawberry (Keniry et al. 2006), Brassica (Burgess et al. 2006; Batley et al. 2007; Hopkins et al. 2007; Ling et al. 2007), Sclerotinia (Winton et al. 2007) and Eragrostis curvula (Cervigni et al. 2008). Analysis of public Brassica data identified 31,155 SSR molecular markers with associated PCR primers. The latest version of the SSRPrimer tool, SSRPrimerII is available at http://acpgg.imb.uq.edu.au/SSRDiscovery.php. Additionally, Brassica SSRPrimer data generated using SSRPRimer are available at http://acpgg.imb.uq.edu.au/brassica.php.

11.2.1.3 SSRPoly

SSRPoly was the first tool to be developed to identify polymorphic SSRs from DNA sequence data. The software consists of a set of Perl scripts and MySQL tables and can be implemented on UNIX, Linux and Windows
platforms. The application of SSRPoly to *Brassica* EST sequences has led to the prediction of 1,528 polymorphic SSRs. The tool can be downloaded from http://acpfg.imb.uq.edu.au/ssrpoly.php.

### 11.2.2 Genetic Mapping

Insight into plant genome organization can be obtained by constructing genetic linkage maps using molecular genetic markers. Molecular markers offer the opportunity to identify the genetic locations of large numbers of regions that govern important agronomic traits, and the resulting molecular genetic maps provide a means to link heritable traits with underlying genome sequence variation. Genetic mapping places markers on linkage groups based on their segregation in a population. Sequence based markers, such as SSRs and SNPs, allow linking of heritable traits with regions of the genome providing a link to the underlying sequence variation associated with observed heritable phenotypic variation. There are several tools available for the generation of genetic maps and the mapping of heritable traits to these maps. These include MapMaker (Lander et al. 1987) and JoinMap (Stam 1993). In addition, there are repositories of *Brassica* genetic mapping data and associated viewers, these include CMAP (http://acpfg.imb.uq.edu.au) and the online *Brassica* repository brassica.info.

#### 11.2.2.1 Brassica.info

Brassica.info is an online repository of *Brassica* information with links to external *Brassica* resources. Resources available from this site include details of genetic maps, markers, sequencing projects and transcriptome information. New data is constantly being collated and updated as it becomes available. This repository can be accessed at http://www.brassica.info.

#### 11.2.2.2 Comparative Mapping

Comparative mapping based on the alignment of chromosomes using common molecular markers helps researchers translate information from one map to another and allows the transfer of knowledge from one genome to another related genome (Chao et al. 1989). Comparative mapping is of particular relevance to the breeding of the *Brassica* crops where conservation between the three *Brassica* genomes permits the transfer of knowledge between these diploids and also with the allopolyploid species.

Restriction fragment length polymorphism (RFLP), SSR and increasingly SNP markers are applied for comparative genetic mapping, since they are often transferable between related species. The linkage arrangement of markers can be compared between closely related species,
if the same molecular markers are used for genetic mapping. This has been demonstrated in *Brassica*, where it has been shown that the linear order of genes is conserved over a large evolutionary timescale between the amphidiploid AB and AC genomes and the diploid progenitor genomes. Comparative genetic mapping may be extended to more divergent species. *Brassica* species are in the same family as *Arabidopsis thaliana*, and these genera diverged approximately 15–21 million years ago (Koch et al. 2000). DNA sequences of homologous genes are similar between the two taxa, and it is, therefore, possible to use markers from one species to map related loci in the other species. Several tools have been developed for comparative genetic mapping, the most common of which is CMap. In addition, the increasing availability of genome sequence data permits the linking of genetic maps with sequenced genomic regions of *Brassica* and related species. This is exemplified by the data within the AtEnsembl database. These tools will become more important as genome sequences become increasingly available.

### 11.2.2.3 CMAP Database

The *Brassica* CMAP database provides comparative genetic information for *Brassica* species. Marker information has been collated from various sources and is presented using a standard graphical interface (Fig. 11-2). The relative locations of markers on *Brassica* chromosomes are presented, and sequence based markers have also been mapped onto the *Arabidopsis* genome. The CMAP database can be accessed through [http://acpfg.imb.uq.edu.au/brassica.php](http://acpfg.imb.uq.edu.au/brassica.php).

### 11.2.2.4 AtEnsembl

AtEnsembl is a database and genome viewer for *A. thaliana*. *Brassica* genomic information in the form of Unigenes and whole genome shotgun data from the *Brassica* C genome has been mapped onto the *Arabidopsis* viewer and can be seen in relation to *Arabidopsis* genomic information including gene expression data and genome insertion mutants. The mapping of *B. rapa* bacterial artificial chromosome (BAC) ends produced as part of the multinational *Brassica* genome sequencing project enables the rapid identification of Chinese cabbage BAC clones predicted to host genes that have been characterized in *Arabidopsis* (Love et al. 2005).
Figure 11-2 The standard graphical interface of the genetic map viewer CMAP showing mapping information for (a) *Brassica rapa* and (b) *Brassica oleracea*. 
11.3 Genomic Resources

11.3.1 Brassica Genome Sequencing

The value of complete genome sequences for species is becoming increasingly recognized. Significant efforts are underway for the full genome sequencing of the important Asian vegetable Chinese cabbage (B. rapa). Currently there are two major Brassica sequencing projects in development as described below.

11.3.1.1 Brassica rapa Genome Sequencing

The Multinational Brassica Genome Sequencing Project aims to sequence the complete genome of Chinese cabbage using a BAC by BAC approach. Three BAC libraries have been produced and end sequences have been obtained for 110,000 BACs. Sequencing is underway for each of the 10 chromosomes and the sequences for several hundred BACs have been submitted to the public GenBank database, including sequences corresponding to the complete Arabidopsis genome. In addition to this project, several groups are undertaking whole genome shotgun sequencing of B. rapa using the Illumina GAII system. The Illumina sequencing is highly complementary to the BAC sequencing project and will assist in the final finishing and polishing of the completed genome sequence.

11.3.1.2 Brassica oleracea Genome Sequencing

In a collaboration between Cold Spring Harbor and TIGR, 0.5× coverage of the B. oleracea genome was produced using whole genome shotgun sequencing. These sequences have been mapped onto the Arabidopsis genome and the results of this comparison can be found in AtEnsEMBL.

11.3.2 Expressed Sequence Tags

Expressed sequence tags (ESTs) correspond to regions of expressed genes. ESTs can be used as a gene discovery tool and also to demonstrate which genes are expressed in a particular tissue at a specific time. Sequencing ESTs is a powerful method of gene discovery and gene sequence determination, and ESTs are also a valuable resource for SNP discovery. There are various resources available that provide information on Brassica ESTs (Table 11-1).
Role of Bioinformatics as a Tool for Vegetable Brassica Species

11.3.2.1 dbEST

dbEST is an online repository of ESTs maintained at the NCBI. It hosts over one million Brassica ESTs and is the primary repository for all EST information. It can be accessed from http://www.ncbi.nlm.nih.gov/dbEST/.

11.3.2.2 UniGene Database at NCBI

A UniGene is derived from a group of ESTs that seem to come from the same gene. The Unigene database currently contains around 50,000 Brassica entries that can be accessed from http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene.

11.3.2.3 CleanEST Database

CleanEST (Lee and Shin 2008) represents a collation of EST libraries that are classified according to species and the sequencing center. It is called a cleansed EST database because it solves the problem of inconsistent terminology and the presence of contaminated sequences. The database currently contains about 500,000 Brassica entries and can be accessed at http://cleanest.kobic.re.kr.

11.3.2.4 AutoSNPdb Clusters

As well as SNP information, AutoSNPdb (Duran et al. 2008) also hosts information on 825,195 ESTs that assemble into 169,721 unigenes consisting of 38,455 contigs and 131,266 singletons.

11.3.3 Gene Expression Information

Along with knowing the gene content of Brassica, it is important to understand which genes are expressed in each of the different tissues at different times and under different environmental conditions. Gene expression microarrays are the favored tool for such high-throughput gene expression analysis. Several independent formats of gene expression microarray have been developed for Brassica species, including Agilent

<table>
<thead>
<tr>
<th>Database</th>
<th>Number of Brassica ESTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbEST at NCBI</td>
<td>~ 1 million</td>
</tr>
<tr>
<td>UniGene database at NCBI</td>
<td>~ 500,000</td>
</tr>
<tr>
<td>Cleaned EST database</td>
<td>~ 500,000</td>
</tr>
<tr>
<td>AutoSNPdb</td>
<td>131,266</td>
</tr>
</tbody>
</table>
Genetics, Genomics and Breeding of Vegetable Brassicas

(http://brassica.bbsrc.ac.uk/array_info.html), Nimblegen (http://www.intl-pag.org/16/abstracts/PAG16_P07a_708.html), Combimatrix (http://www.brassica-rapa.org/BGP/NC_DNA_chip.jsp) and spotted cDNA arrays (http://www.intl-pag.org/14/abstracts/PAG14_P740.html). In addition to Brassica specific arrays, the Arabidopsis Affymetrix GeneChip has also been successfully applied for Brassica transcriptomics (http://affymetrix.arabidopsis.info/xspecies/). It is expected that with the increased availability of second generation sequencing technology this would be applied for transcriptomics in Brassica. However there are currently no published reports of this application.

11.4 Future Prospects

There are many tools and information resources available for genetic and genomic analysis in Brassica. However, the everchanging technologies will demand increasingly complex and powerful systems to translate the massive amount of data being generated to improve Brassica vegetable crops. The principle challenges and opportunities come from the application of new sequencing technologies. One may predict that in the near future, researchers will be able to sequence the genome of varieties or wild relatives to identify the underlying basis of phenotypic variation, and then use this information directly for crop improvement. This will require changes in the way we manage Brassica data, particularly phenotypic data, to ensure that robust associations are made between the emerging genome sequences and valuable agronomic traits.

References


Role of Bioinformatics as a Tool for Vegetable Brassica Species


