Plant Genetics and Genomics: Crops and Models 9

Renate Schmidt
Ian Bancroft
Editors

Genetics and Genomics of the Brassicaceae

Springer

Plant Genetics and Genomics: Crops and Models

Volume 9

Series Editor
Richard A. Jorgensen

For further volumes:
http://www.springer.com/series/7397

Renate Schmidt · Ian Bancroft
Editors

Genetics and Genomics of the Brassicaceae

Springer

Editors
Renate Schmidt
Leibniz-Institut für
Pflanzengenetik und
Kulturpflanzenforschung (IPK)
Corrensstr.3
06466 Gatersleben
Germany
schmidtr@ipk-gatersleben.de

Ian Bancroft
John Innes Institute
Norwich Research Park
NR4 7UH Colney, Norwich
United Kingdom
ian.bancroft@bbsrc.ac.uk

ISBN 978-1-4419-7117-3 e-ISBN 978-1-4419-7118-0
DOI 10.1007/978-1-4419-7118-0
Springer New York Dordrecht Heidelberg London

Preface

The angiosperm family Brassicaceae is commonly termed the mustard family or, because of their characteristic flowers consisting of four petals in the form of a Greek cross, the Cruciferae. It comprises 338 genera assigned to 25 tribes and includes the widely studied species *Arabidopsis thaliana* (L.) Heynh. (thale cress) of the Camelineae and a diverse array of cultivated types within the Brassiceae, including oilseed rape, mustards, leafy vegetables, root vegetables, and cole (stem vegetable) crops.

The initial focus for the application of genomic approaches in the Brassicaceae was *A. thaliana*, which had been selected by the late 1980s as a "model" species in which to study plant biology at the molecular level. This was the first plant species for which a genome sequencing program was launched in the mid-1990s by the *Arabidopsis* Genome Initiative, culminating in 2000 in the landmark publication of an analysis of its complete genome sequence. This resource has facilitated the unprecedented expansion in our understanding of plant biology over the last decade.

Genomic and comparative genomic analyses have been applied to a number of species within the Brassicaceae, revealing much about genome evolution in plants, particularly after the publication of *Arabidopsis* genome sequence. Such studies revealed, for example, that the ancestral karyotype for the Brassicaceae was probably $n=8$, and that numerous chromosomal rearrangements and a reduction in chromosome number shaped the genome of *A. thaliana*. Angiosperms have a propensity to undergo chromosome doubling, or polyploidization. Such events are followed by a process of "diploidization," during which genomes stabilize and gene copy number is reduced. The Brassicaceae presents an excellent opportunity to study these processes. The genome sequence of *A. thaliana* provides evidence for as many as three polyploidization events, the last of which occurred near the origin of the Brassicaceae, and is anticipated to be present throughout the family. A distinctive feature of the tribe Brassiceae is extensive subsequent genome triplication, indicative of a hexaploidy event. In addition, several species, particularly within the Brassiceae, are recently formed allotetraploids, e.g., *Brassica napus* ($n=19$) was formed by hybridization of *Brassica rapa* ($n=10$) and *B. oleracea* ($n=9$). Our increasing understanding of these processes is crucial for the interpretation of data from comparative genomic analyses.

The most extensive genomic resources have been developed for the tribe Camelineae, principally, though not exclusively, for *A. thaliana*. These include 130 Mb genome sequence of *A. thaliana*, 1.5 m *A. thaliana* ESTs, multiple commercially available *A. thaliana* microarrays, publicly available gene knockout lines, resource centers for plant lines and DNA stocks, and public databases.

Genomic resources are rapidly developing for the tribe Brassiceae, mainly driven by the economic importance of the *Brassica* species crops. These include ongoing genome sequencing of the *B. rapa* genespace, 0.8 m ESTs from *Brassica* species, a commercially available *Brassica* microarray, many linkage maps (some being integrated), mutagenized populations, resource centers for distribution of BAC libraries, public databases, and a steering committee to coordinate research efforts.

Genomic resources are being developed for a few species from other tribes, which are being studied primarily because of particular attributes, e.g., *Thlaspi caerulescens* which is being studied for metal hyperaccumulation, *Arabis alpina* for perenniality, *Thellungiella halophila* for salt tolerance, and *Boechera* sp. for apomixis.

Contents

Contributors

Carlos Alonso-Blanco Departamento de Genética Molecular de Plantas, Centro Nacional de Biotecnologia (CNB) Consejo Superior de Investigaciones Científicas (CSIC), 28049 Madrid, Spain, calonso@cnb.csic.es

Ian Bancroft John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK, ian.bancroft@bbsrc.ac.uk

Christoph Böttcher Department of Stress and Developmental Biology, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle/Saale, Germany, cboettch@ipb-halle.de

Deborah Crist Department of Plant Cellular and Molecular Biology, Arabidopsis Biological Resource Center, The Plant Biotechnology Center, The Ohio State University, Columbus, OH 43210, USA, abrc@arabidopsis.org

Mark W. Farnham USDA-ARS-U.S. Vegetable Laboratory, Charleston, SC 29414, USA, mark.farnham@ARS.USDA.GOV

Maria L. Federico Genomics and Bioinformatics Unit (UGB), Agri aquaculture Nutritional Genomic Center (CGNA), INIA-Carillanca Km 10 Camino Cajon-Vilcun, PO Box 58D, Temuco, Chile, maria.federico@cgna.cl

Robert T. Gaeta Division of Biological Sciences, University of Missouri, Columbia, MO 65211, USA, gaetar@missouri.edu

Cassandra M.P. Goldsack John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK, cassie.goldsack@bbsrc.ac.uk

Neil Graham Plant and Crop Sciences Division, Nottingham Arabidopsis Stock Centre, University of Nottingham, Loughborough, LE12 5RD UK, neil@arabidopsis.org.uk

Erich Grotewold Department of Plant Cellular and Molecular Biology, Arabidopsis Biological Resource Center, The Plant Biotechnology Center, The Ohio State University, Columbus, OH 43210, USA, abrc@arabidopsis.org

Georg Haberer MIPS, Institute for Bioinformatics and Systems Biology, Helmholtz Center Munich, 85764 Neuherberg, Germany, g.haberer@helmholtz-muenchen.de

Federico L. Iniguez-Luy Genomics and Bioinformatics Unit (UGB), Agri aquaculture Nutritional Genomic Center (CGNA), INIA-Carillanca Km 10 Camino Cajon-Vilcun, PO Box 58D, Temuco, Chile, federico.iniguez@cgna.cl

Emma M. Knee Department of Plant Cellular and Molecular Biology, Arabidopsis Biological Resource Center, The Plant Biotechnology Center, The Ohio State University, Columbus, OH 43210, USA, abrc@arabidopsis.org

Marcus A. Koch Heidelberg Institute of Plant Sciences, Biodiversity and Plant Systematics, University of Heidelberg, Heidelberg, Germany, mkoch@hip.uni-hd.de

Maarten Koornneef Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany, koornnee@mpipz.mpg.de

Helmi Kuittinen Department of Biology and Biocenter, University of Oulu, Oulu FIN-90014, Finland, helmi.kuittinen@oulu.fi

Yong Pyo Lim Department of Horticulture, Chungnam National University, Daejeon 305-764, South Korea, yplim@cnu.ac.kr

Martin A. Lysak Department of Functional Genomics and Proteomics, Institute of Experimental Biology, Masaryk University, Brno, Czech Republic, lysak@sci.muni.cz

Sean May Plant and Crop Sciences Division, Nottingham Arabidopsis Stock Centre, University of Nottingham, Loughborough, LE12 5RD UK, sean@arabidopsis.org.uk

Klaus F.X. Mayer MIPS, Institute for Bioinformatics and Systems Biology, Helmholtz Center Munich, 85764 Neuherberg, Germany, k.mayer@helmholtz-muenchen.de

Jeong-Hwan Mun Genomics Division, Department of Agricultural Biotechnology, National Academy of Agricultural Science, Suwon 441-707, Korea, munjh@rda.go.kr

June B. Nasrallah Department of Plant Biology, Cornell University, Ithaca, NY 4853, USA, jbn2@cornell.edu

Lars Østergaard John Innes Centre, Norwich, UK, lars.ostergaard@bbsrc.ac.uk

Beom-Seok Park Genomics Division, Department of Agricultural Biotechnology, National Academy of Agricultural Science, Suwon 441-707, Korea, pbeom@korea.kr

Isobel Parkin Agriculture and Agri-Food Canada, Saskatoon Research Centre, Saskatoon, SK, Canada S7N 0X2, isobel.parkin@agr.gc.ca

Deepak Pental Department of Genetics, Centre for Genetic Manipulation of Crop Plants, University of Delhi South Campus, New Delhi 110021, India, dpental@gmail.com

J. Chris Pires Division of Biological Sciences, 371 Life Sciences Center, University of Missouri, Columbia, MO 65211, USA, piresjc@missouri.edu

Akshay K. Pradhan Department of Genetics, Centre for Genetic Manipulation of Crop Plants, University of Delhi South Campus, New Delhi 110021, India, pradhancgmcp@gmail.com

Carlos F. Quiros Department of Plant Sciences, University of California, Davis, CA 95616, USA, cfquiros@ucdavis.edu

Nirala Ramchiary Department of Horticulture, Chungnam National University, Daejeon 305-764, South Korea, nrudsc@gmail.com

Matthieu Reymond Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany; Institut Jean-Pierre Bourgin, UMR1318 INRA-AgroParisTech, INRA Centre de Versailles-Grignon, Route de St-Cyr (RD10), 78026 Versailles Cedex, France, reymond@mpipz.mpg.de

Luz Rivero Department of Plant Cellular and Molecular Biology, Arabidopsis Biological Resource Center, The Plant Biotechnology Center, The Ohio State University, Columbus, OH 43210, USA, abrc@arabidopsis.org

Outi Savolainen Department of Biology and Biocenter, University of Oulu, Oulu FIN-90014, Finland, outi.savolainen@oulu.fi

Dierk Scheel Department of Stress and Developmental Biology, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle/Saale, Germany, dscheel@ipb-halle.de

Renate Schmidt Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), 06466 Gatersleben, Germany, schmidtr@ipk-gatersleben.de

Randy Scholl Department of Plant Cellular and Molecular Biology, Arabidopsis Biological Resource Center, The Plant Biotechnology Center, The Ohio State University, Columbus, OH 43210, USA, scholl.1@osu.edu; abrc@arabidopsis.org

Penny A.C. Sparrow John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK, penelope.sparrow@bbsrc.ac.uk

Pauline Stephenson John Innes Centre, Norwich, UK, pauline.stephenson@bbsrc.ac.uk

Günter Theißen Department of Genetics, Friedrich Schiller University Jena, Jena D-07743, Germany, guenter.theissen@uni-jena.de

Chris Town J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA, cdtown@jcvi.org

Martin Trick Computational and Systems Biology Department, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK, martin.trick@bbsrc.ac.uk

Bekir Ülker Plant Molecular Engineering Group, IZMB (Institute of Cellular and Molecular Botany), University of Bonn, D-53115 Bonn, Germany, ulker@uni-bonn.de

Edda von Roepenack-Lahaye Department of Stress and Developmental Biology, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle/Saale, Germany, vonroep@ipb-halle.de

Yu Wang MIPS, Institute for Bioinformatics and Systems Biology, Helmholtz Center Munich, 85764 Neuherberg, Germany, yu.wang@helmholtz-muenchen.de

Suzanne I. Warwick Eastern Cereal and Oilseed Research Centre, Agriculture and Agri-Food Canada, Central Experimental Farm, Ottawa, ON, Canada K1A OC6, warwicks@agr.gc.ca

Bernd Weisshaar Faculty of Biology, Bielefeld University, D-33594, Bielefeld, Germany, bernd.weisshaar@uni-bielefeld.de

Thomas Wood John Innes Centre, Norwich, UK, tom.wood@bbsrc.ac.uk

Chapter 1
Phylogeny, Genome, and Karyotype Evolution of Crucifers (Brassicaceae)

Martin A. Lysak and Marcus A. Koch

Abstract Brassicaceae (crucifers or the mustard family) is a large plant family comprising over 330 genera and about 3,700 species, including several important crop plants (e.g. *Brassica* species), ornamentals as well as model organisms in the plant sciences (e.g. *Arabidopsis thaliana*). In recent years, the wealth of *Arabidopsis* and *Brassica* genomic resources along with newly established tools and techniques fostered the unprecedented progress in phylogenetics and genomics of crucifers. Multigene phylogenetic analyses paved the way for a new infrafamiliar classification based on phylogenetically circumscribed genera and tribes. A new generation of comparative genetic, cytogenetic, and genomic studies as well as whole-genome sequencing projects unveil general principles of karyotype and genome evolution in Brassicaceae.

Keywords Brassicaceae · Cruciferae · Phylogeny · Tribal classification · Genome and karyotype evolution · Chromosomes · Genome size · Whole-genome duplication · Polyploidy · Genome collinearity

Abbreviations

ACK	Ancestral crucifer karyotype
AK	Ancestral chromosome of the ACK
ancGS	Ancestral genome size
APG	Angiosperm phylogeny group
CCP	Comparative chromosome painting
cpDNA	Chloroplast DNA
DAPI	4′, 6-diamidino-2-phenylindole
FISH	Fluorescence in situ hybridization
gDNA	Genomic DNA
GISH	Genomic in situ hybridization

M.A. Lysak (✉)
Department of Functional Genomics and Proteomics, Institute of Experimental Biology, Masaryk University, Brno, Czech Republic
e-mail: lysak@sci.muni.cz

R. Schmidt, I. Bancroft (eds.), *Genetics and Genomics of the Brassicaceae*, Plant Genetics and Genomics: Crops and Models 9, DOI 10.1007/978-1-4419-7118-0_1,

GS	Genome size
ITS	Internal transcribed spacer within rDNA
NOR	Nucleolar organizing region
mtDNA	Mitochondrial DNA
mya	Million years ago
PCK	Proto-Calepineae karyotype
rDNA	Ribosomal DNA
RFLP	Restriction fragment length polymorphism
RGC	Rare genomic change
WGD	Whole-genome duplication

Contents

1.1 General Introduction

The Brassicaceae (Cruciferae or mustard family) is a large plant family with approximately 338 genera and 3,709 species (see Warwick et al. 2006b) and is of special interest as it include many crop plants (*Brassica oleracea*, *Brassica napus*, *Armoracia rusticana*, and many more), ornamentals (*Aubrieta*, *Iberis*, *Lunaria*, *Arabis*, *Draba* and others) as well as model organisms in the plant sciences (*Arabidopsis thaliana*, *Arabidopsis lyrata*, *Arabidopsis halleri*, *B. napus*, *Capsella rubella*, *Thellungiella halophila*, *Arabis alpina*, and few others). The family shows a worldwide distribution, except Antarctica (Fig. 1.1). Most of the taxa are found in temperate regions of the Northern Hemisphere. However, numerous genera are also found in the Southern Hemisphere (such as *Draba*, *Lepidium*, and *Cardamine*), and some of them are even endemic to southern regions (e.g., South African genera:

Fig. 1.1 Worldwide distribution of the Brassicaceae

Brachycarpea, *Chamira*, *Schlechteria, Silicularia*, all subsumed by Al-Shehbaz and Mummenhoff 2005 under a broadly defined genus *Heliophila*). In the tropics, the distribution of Brassicaceae is limited to mountainous and alpine regions. *A. alpina* represents the classical example of a plant with a worldwide Northern Hemispheric distribution in mountainous, alpine and arctic habitats including the East African high mountains in Kenya, Tanzania and Ethiopia (Koch et al. 2006). The worldwide distribution of Brassicaceae provides an excellent basis to perform various evolutionary, biogeographic or phylogeographic studies on various taxonomic levels (Koch and Kiefer 2006). However, species diversity is not distributed equally and the most important diversification centers are found in the Irano-Turanian region (ca. 150 genera and ca. 900 species with 530 endemics) and the Mediterranean region (ca. 113 genera and ca. 630 species with 290 endemics). The Saharo-Sindian region (ca. 65 genera and 180 species with 62 endemics) and North America (ca. 99 genera and 778 species with 600 endemics) show a significant reduction in species diversity (Hedge 1976, Al-Shehbaz 1984, Rollins 1993, Appel and Al-Shehbaz 2002). This reduction of species diversity is continued in the Southern Hemisphere (South America with 40 genera and 340 species; Southern Africa with 15 genera and at least 100 species; and Australia and New Zealand with 19 genera and 114 species) (Allan 1961, Marais 1970, Hewson 1982, Al-Shehbaz 1984, Appel and Al-Shehbaz 2002). This overall distribution pattern might provide some evidence for the origin of the family in the Irano-Turanian region (Franzke et al. 2009). It should be noted that the taxonomic circumscription of many taxa, on the generic as well as species level, is still provisional. This is obvious, when we have a closer look on the number of species within a given genus (Fig. 1.2) with 62 or 72% of the genera with three species or less and five species or less, respectively.

Species of Brassicaceae can be easily distinguished morphologically from species of any other family of vascular plants, even if closely related, due to its highly conserved and fairly uniform flower architecture. The flowers are bilaterally symmetrical and consist of four, almost always free sepals in two whorls, free petals (though sometimes lacking), often six, free, tetradynamous stamens (outer

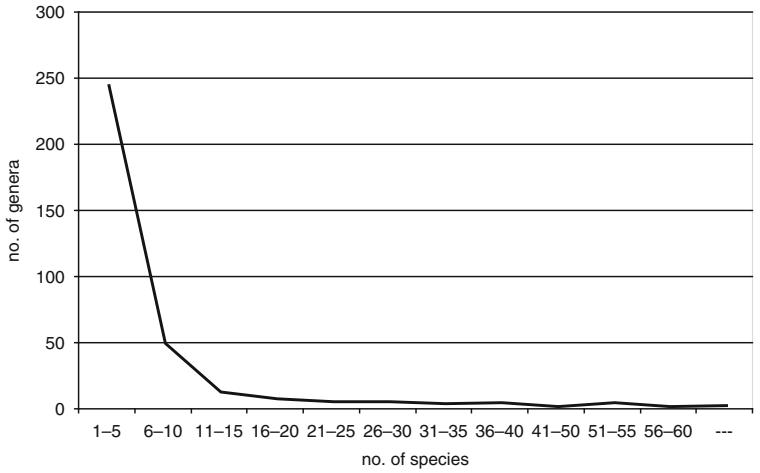

Fig. 1.2 Distribution of species number within crucifer genera

two shorter than inner four; though many *Lepidium* species have four or only two stamens), and a bicarpellate ovary. Only few species and genera show an asymmetrical perianth (e.g., *Iberis* or *Teesdalia*) or divided petals (*Berteroa*). In contrast, fruits of the Brassicaceae exhibit enormous diversity in size, shape, and structure. Therefore, the fruits are used as the most important diagnostic characters for the delimitation and identification of taxa at various ranks. The fruits are often dehiscent, two-valved capsules divided longitudinally by a false septum into two loculi, though in many genera the fruits are indehiscent and/or the septum is totally lacking. Some groups are characterized by angustiseptate fruits (compressed at a right angle to the septum), as in the members of the tribe Lepidieae sensu Schulz (1936), while in the majority of the family the fruits are either latiseptate (compressed parallel to the septum) or not flattened (terete or angled). On the basis of length/width ratio, the fruits have been traditionally divided into those with silicles (length less than three times the width) or with siliques (length more than three times the width), but such division, though useful diagnostically, is arbitrary and has no phylogenetic implications (Al-Shehbaz 1984, Appel and Al-Shehbaz 2002, Al-Shehbaz et al. 2006). Additional important taxonomic characters include embryologic features (position of cotyledons in relation to radicle), nectary glands, trichomes, as well as growth form, chromosome numbers, and seed-coat anatomy and surface (Appel and Al-Shehbaz 2002).

The history of systematics, phylogenetics, and evolutionary research in Brassicaceae family can be divided into three periods. The first period provided us with comprehensive taxon descriptions and several classification systems proposed from the early nineteenth to the mid-twentieth century. The most notable among them are those of De Candolle (1821), Prantl (1891), Hayek (1911), Schulz

(1936), and Janchen (1942). According to these systems, Brassicaceae was artificially divided into 4–19 tribes and 20–30 subtribes. A second period cumulated after a summary of knowledge was obtained about this family more than 30 years ago (Vaughan et al. 1976), followed by a contribution of Tsunoda et al. (1980) that dealt with the biology and breeding of *Brassica* crops and their wild allies. This period was also characterized with a series of papers describing species and their taxonomy and re-defining various tribes and subtribes (e.g. Al-Shehbaz 1984, 1985, 1988a, b, c, for review see also Koch 2003, Koch et al. 2003). The third and most dynamic period of significant taxonomic changes started in the early 1990s, initially based on isozymes but continuously increasing the amount of utilized DNA data (e.g. Koch 2003, Koch et al. 2003). Since then molecular biology and DNA techniques have revolutionized plant systematics and phylogenetics, and because of the selection of *A. thaliana* (*Arabidopsis* thereafter) as the most prominent model plant system, crucifer species belong to an intensely studied plant group.

What are the most important milestone accomplishments in crucifer phylogenetics during the past two decades? In principle, and aside from the wealth of knowledge and resources for *A. thaliana* and other model species, they are at least four: (i) achieving a new infrafamiliar classification based on phylogenetically circumscribed genera and tribes, (ii) phylogenetic circumscription of the order Brassicales and the determination of Cleomaceae as the closest and sister family to Brassicaceae, (iii) unravelling general principles of crucifer evolution by exploring species- or genus-specific evolutionary histories, and (iv) detailed information on karyotype and genome evolution across the family including genetic and cytogenetic maps as well as whole-genome DNA sequence data sets. In this chapter, these four principal achievements were divided into two main sections: Phylogenetic position of Brassicaceae and recognition of infrafamiliar taxa (Section 1.2) and genome and chromosomal evolution (Section 1.3).

1.2 Phylogenetic Position of Brassicaceae and Recognition of Infrafamiliar Taxa

In the most recent comprehensive angiosperm phylogeny (Stevens 2001) the order Brassicales (extended order Capparales) comprises 17 families, 398 genera, and approximately 4,450 species. Overall Brassicales contains approximately 2.2% of the eudicot diversity (Magallön et al. 1999) with its earliest fossil known from the Turonian [89.5 million years ago (mya)]. The age of a stem and crown group was estimated to be 90–85 and 71–69 mya, respectively (Wikström et al. 2001).

Based on strictly morphological studies, Judd et al. (1994) indicated that Brassicaceae is nested within the paraphyletic Capparaceae (including Cleomaceae) and suggested their union into one family, Brassicaceae s.l. However, molecular studies (Hall et al. 2002, 2004, Schranz and Mitchell-Olds 2006) clearly demonstrated that Brassicaceae is sister to Cleomaceae and both are sister to Capparaceae. As a result, three families are currently recognized.

The history of tribal classification systems is long and well summarized in various reviews (e.g., Appel and Al-Shehbaz 2002, Koch 2003, Koch et al. 2003, Mitchell-Olds et al. 2005, Al-Shehbaz et al. 2006) and does not need to be repeated here. Prior to 2005 as summarized in Koch and Mummenhoff (2006), the most important conclusion reached in phylogenetic studies was that except for the Brassiceae, the other tribes were artificially delimited and did not reflect the phylogenetic relationships of their component genera. The other exception was thought to be the tribe Lepidieae (e.g., Zunk et al. 1999), but that too was shown to be artificially circumscribed (Al-Shehbaz et al. 2006).

Of the 49 infrafamiliar taxa (19 tribes and 30 subtribes) previously recognized by Schulz (1936), only nine tribes (Alysseae, Arabideae, Brassiceae, Euclidieae, Heliophileae, Hesperideae, Lepidieae, Schizopetaleae, and Sisymbrieae) were maintained in the re-circumscription of Brassicaceae by Al-Shehbaz et al. (2006). In addition, 16 tribes were either newly described or re-established. This tribal classification followed the first comprehensive phylogeny of Brassicaceae based on the plastidic gene *ndh*F (Beilstein et al. 2006). The study identified three major, significantly supported clades (Lineage I–III; Fig. 1.3). The authors have recently confirmed these data with an extended data set utilizing also nuclear phytochrome A sequence data (Beilstein et al. 2008). A subsequent, ITS-based phylogeny (Bailey et al. 2006) provided substantial support to the new tribal system. In a more recent analysis focusing primarily on the evolution of plastidic *trn*F pseudogene in the mustard family, a supernetwork was reconstructed based on nuclear alcohol dehydrogenase (*adh*), chalcone synthase (*chs*), internal transcribed spacer of nuclear ribosomal DNA (ITS), and plastidic maturase (*mat*K) sequence data (Koch et al. 2007). The supertree was largely in congruence with the corresponding *trn*LF derived phylogeny, and all three major lineages identified by Beilstein et al. (2006) were confirmed. The supertree approach clearly demonstrated that there is a substantial conflicting "phylogenetic signal" at the deeper nodes of the family tree resulting in virtually unresolved phylogenetic trees at the generic level. Some results were contradictory, such as the ancestral position of the Cochlearieae (Koch et al. 2007), which was not revealed by the *ndh*F (Beilstein et al. 2006, 2008) or ITS data (Bailey et al. 2006). Remarkably, a phylogenetic study using the mitochondrial *nad*4 intron (Franzke et al. 2009) is highly congruent with the ITS- and *ndh*F-based studies. In summary, most of the tribes recognized by Al-Shehbaz et al. (2006) are clearly delimited, however, much less significant support is available for the relationships between the various tribes.

Despite the use of multi-gene phylogenies, the lack of resolution in the skeletal backbone of the family's phylogenetic tree is not yet understood, which could be due to one of the following hypotheses. First, early radiation events were quite rapid and were characterized by low levels of genetic variation separating the different lineages. Second, reticulate evolution (e.g., as found in the tribe Brassiceae) resulted in conflicting gene trees that do not reflect species phylogenies. The mitochondrial *nad*4 intron data presented by Franzke et al. (2009) perhaps favor the former hypothesis. Koch et al. (2007), who found that the micro-structural evolutionary changes may be useful for inferring early events of divergence, also

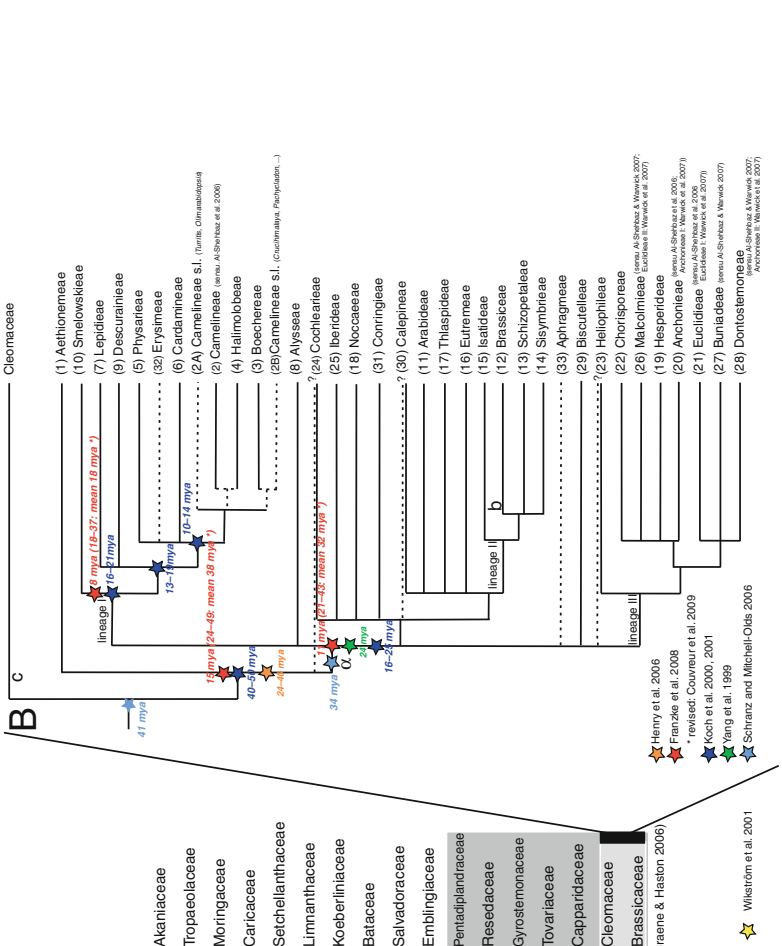

Fig. 1.3 Synopsis of phylogenetic hypothesis, divergence time estimates, and genome duplication events (see also Fig. 1.4). **a** Brassicales according to APG II (2003). **b** Brassicaceae: combination from various sources of tribal relationships in the Brassicaceae (for details refer to the text). Lineages I–III are described in Beilstein et al. (2006), (2008). Koch et al. (2007) used different numbers and we suggest to use Beilstein's version only in order to avoid future confusion. *Dashed lines* indicate uncertain phylogenetic position. However, it should be kept in mind that this synopsis is not derived from one single phylogenetic analysis. All branches are not drawn to scale

favored this scenario. In fact, the two structural rearrangements in the *trn*L-F region alone (Koch et al. 2007) indicate ancient patterns of divergence and their early occurrence is supported by phylogenetic analysis of that region even while excluding the micro-structural mutations (Koch et al. 2007) and is further supported from analyses of the nuclear ITS sequence data (Bailey et al. 2006).

The recently proposed tribal classification of Al-Shehbaz et al. (2006) recognized 25 tribes (see below). More recently, Franzke et al. (2009) presented a family phylogeny based on the mitochondrial *nad*4 intron. Although the sampling in the latter study was smaller, both cpDNA (Beilstein et al. 2006, 2008) and mtDNA (Franzke et al. 2009) phylogenies are mostly in congruence with each other. However, it is still unclear why there are major inconsistencies between the two phylogenies and those based on nuclear markers, such as the ITS (Bailey et al. 2006), *adh*, and *chs* (Koch et al. 2000, 2001).

Additional studies have shown that some of the tribes proposed by Al-Shehbaz et al. (2006) were broadly delimited or are paraphyletic and need further splitting. For example, Warwick et al. (2007) and Al-Shehbaz and Warwick (2007) showed the Euclidieae and Anchonieae to consist of more than one lineage and were newly defined as the Malcolmieae, Dontostemoneae, and the re-established Buniadeae (tribes 26–28, Fig. 1.3). German and Al-Shehbaz (2008) described the new Aphragmeae and Conringieae and re-established Calepineae, Biscutelleae, and Erysimeae (tribes 29–33, Fig. 1.3). ITS studies of Bailey et al. (2006) and Warwick et al. (2010) justify the recognition of the last tribe. They also demonstrated that the Camelineae sensu Al-Shehbaz et al. (2006) is paraphyletic and requires further division into at least three different groups, herein recognized as tribes 2, 34 (2A), and 35 (2B) (Table 1.1, Fig. 1.3). An overview of these various tribes and a synopsis of their relationships are presented in Fig. 1.3. However, this figure does not represent an outcome of a comprehensive family-wide phylogenetic study which is still lacking. Furthermore, it should be emphasized that phylogenetic hypotheses based on single markers (e.g., plastidic, mitochondrial or nuclear) possess a limited value (Koch et al. 2001, 2007). In order to establish a comprehensive phylogeny of the entire family, several taxonomically problematic genera need to be sampled and adequately assigned to tribes (Al-Shehbaz et al. 2006). For further generic delimitations, the interested reader should consult Appel and Al-Shehbaz (2002) and the database of Warwick et al. (2006b); the most comprehensive modern tribal classification was given by Al-Shehbaz et al. (2006).

Although a complete tribal classification system of Brassicaceae is not yet available, we are gradually approaching this goal. Following the first phylogenetic tribal classification of the family (Al-Shehbaz et al. 2006), subsequent molecular studies (e.g., Bailey et al. 2006, Beilstein et al. 2006, 2008, Warwick et al. 2006a, 2007, 2008, Koch et al. 2007, Koch and Al-Shehbaz 2009) led to the tribal adjustments recently proposed by Al-Shehbaz and Warwick (2007) and German and Al-Shehbaz (2008). Table 1.1 summarizes and updates our knowledge of the tribal placement of nearly two-thirds (62.7%) of the 338 genera and 87.6% of the 3,709 species compiled by Warwick et al. (2006b).

Table 1.1 Overview on tribes, genera, and species of Brassicaceae analyzed for their phylogenetic position using molecular markers (Note: The family comprises approximately 338 genera and 3,709 species in total, Warwick et al. 2006b)

	No. of Genera	No. of Species	References
1. Aethionemeae	1	45	Koch and Al-Shehbaz (2009)
2. Camelineae	7	35	Koch and Al-Shehbaz (2009)
3. Boechereae	7	118	Al-Shehbaz et al. (2006)
4. Halimolobeae	5	39	Bailey et al. (2007)
5. Physarieae	7	133	Koch and Al-Shehbaz (2009)
6. Cardamineae	9	333	Koch and Al-Shehbaz (2009)
7. Lepidieae	4	235	Koch and Al-Shehbaz (2009)
8. Alysseae	15	283	Koch and Al-Shehbaz (2009), Warwick et al. (2008)
9. Desurainieae	6	57	Al-Shehbaz et al. (2006)
10. Smelowskieae	1	25	Al-Shehbaz et al. (2006)
11. Arabideae	8	470	Koch and Al-Shehbaz (2009)
12. Brassiceae	46	230	Al-Shehbaz et al. (2006)
13. Schizopetaleae s.l.	28	230	Al-Shehbaz et al. (2006)
14. Sisymbrieae	1	40	Al-Shehbaz et al. (2006)
15. Isatideae	2	65	Koch and Al-Shehbaz (2009)
16. Eutremeae	1	26	Warwick and Al-Shehbaz (2006)
17. Thlaspideae	7	27	Al-Shehbaz et al. (2006)
18. Noccaeeae	3	90	Koch and Al-Shehbaz (2009)
19. Hesperideae	1	45	Al-Shehbaz et al. (2006)
20. Anchonieae	8	68	Al-Shehbaz and Warwick (2007)
21. Euclidieae	13	115	Al-Shehbaz and Warwick (2007)
22. Chorisporeae	3	47	Al-Shehbaz and Warwick (2007)
23. Heliophileae	1	80	Al-Shehbaz et al. (2006)
24. Cochlearieae	1	21	Al-Shehbaz et al. (2006)
25. Iberideae	1	27	Al-Shehbaz et al. (2006)
26. Malcolmieae	8	37	Al-Shehbaz and Warwick (2007)
27. Buniadeae	1	3	Al-Shehbaz and Warwick (2007)
28. Dontostemoneae	3	28	Al-Shehbaz and Warwick (2007)
29. Biscutelleae	1	53	German and Al-Shehbaz (2008)
30. Calepineae	3	8	German and Al-Shehbaz (2008)
31. Conringieae	2	9	German and Al-Shehbaz (2008)
32. Erysimeae	1	180	German and Al-Shehbaz (2008)
33. Aphragmeae	1	11	German and Al-Shehbaz (2008)
34. Unnamed (Camelineae 2A)	2	5	Koch and Al-Shehbaz (2009)
35. Unnamed (Camelineae 2B)	3	20	Koch and Al-Shehbaz (2009)
Total	212	3,249	

The following paper added a significant number of unassigned genera to existing datasets and as a consequence bringing the total number of tribes to 44 [WARWICK S.I., MUMMENHOFF K., SAUDER C.A., KOCH M.A., AL-SHEBAZ I.A. (2010) Closing the gaps: Phylogenetic relationships in the Brassicaceae based on DNA sequence data of nuclear ribosomal ITS region. Pl. Syst. Evol. 285 (3-4): 209-232.]

An ongoing comprehensive phylogenetic study of the family on the generic level (Warwick et al. 2010) aims to cover more than 95% of all currently recognized genera. The major difficulty is obtaining adequate material for molecular studies on species of numerous monospecific or oligospecific genera (see Fig. 1.2). Many of these are known only from the collections of their type species. As most of the larger crucifer genera (e.g., *Cardamine, Draba*, *Erysimum, Heliophila, Lepidium, Rorippa*) are reasonably well-surveyed molecularly and were shown to be largely monophyletic, it is the smaller and medium-sized genera (especially of the tribes Brassiceae and Schizopetaleae) that need further studies. We assume that many of these genera will be merged with others, and that the total number of genera might be reduced substantially.

1.3 Genome and Chromosomal Evolution

1.3.1 Prehistory of Crucifer Genomes: Whole-Genome Duplications and the Age of the Family

One of the fundamental recent findings of crucifer genomics is the importance of ancient whole-genome duplications (WGDs) which influenced literally every structural and functional aspect of Brassicaceae genomes. The sequencing of the *Arabidopsis* genome revealed that at least 60% of its genome is segmentally duplicated (AGI 2000). Over the past several years a general consensus has been reached on the origin of duplicated segments in *Arabidopsis* through ancient tetraploidization events α, β, and γ (Blanc et al. 2003, Bowers et al. 2003, De Bodt et al. 2005, Henry et al. 2006, Maere et al. 2005, Simillion et al. 2002). Nevertheless, the concept of the three WGDs has been challenged by the analysis of whole-genome shotgun sequence of grapevine (*Vitis vinifera*). Jaillon et al. (2007) showed that the grapevine genome, consisting of three ancestral genomes, has a paleohexaploid origin. The comparison with other plant genome sequences available (*Arabidopsis*, poplar, and rice) showed that the paleohexaploidy event must have occurred before the radiation of Eurosids and probably after the separation of dicots and monocots. In *Arabidopsis*, the ancient triplication was equated with the WGD event γ, which was followed by subsequent tetraploidization events β and α (Jaillon et al. 2007). As the two latter paleotetraploidy events were not identified in Caricaceae (papaya, *Carica papaya*; Ming et al. 2008, Tang et al. 2008), belonging to the Brassicales as *Arabidopsis*, it can be assumed that the two paleopolyploid events are specific for the core Brassicales including Brassicaceae (cf. Hall et al. 2004). Furthermore, an independent paleohexaploid WGD detected in Cleomaceae (Schranz and Mitchell-Olds 2006), a sister family of Brassicaceae, is suggesting that the α WGD event post-dates the Cleomaceae–Brassicaceae split. In addition, the absence of two paralogs of arginine decarboxylase gene (*Adc*) in the genus *Aethionema* (Aethionemeae) as compared to the remaining Brassicaceae species (Galloway et al. 1998, Schranz and Mitchell-Olds 2006) pinpoints the α event to be specific for "core Brassicaceae." Further research is needed to resolve

the position of β and α WGDs within Brassicales and Brassicaceae, respectively. The position of the four WGDs within a phylogenetic framework and their dating are shown in Fig. 1.4a, b, respectively.

Comparative genetic mapping using *Arabidopsis* markers has shown that genomes of *Brassica* species with diploid-like chromosome numbers were triplicated due to a hexaploidy event (Lagercrantz 1998, Parkin et al. 2005). The whole-genome triplication has also been unambiguously evidenced cytogenetically by comparative chromosome painting (CCP) using *Arabidopsis* BAC contigs in several *Brassica* species and the tribe Brassiceae (Lysak et al. 2005, 2007, Ziolkowski et al. 2006). Therefore, such a relatively recent genome duplication which can be uncovered by CCP and comparative genetic mapping, unlike the three ancient WGDs, is classified here as a mesopolyploid WGD event. Ongoing CCP analyses in a wide range of Brassicaceae species indicate that mesopolyploid WGDs were relatively common within different crucifer lineages (Mandáková et al. 2010 and unpublished data). Paleopolyploid and mesopolyploid WGDs were obscured by chromosomal and genetic diploidization processes including chromosome rearrangements, leading to chromosome number reduction, genome size downsizing, and diploid-like inheritance. Diploid-like genomes give rise to neopolyploid cytotypes and species via autopolyploidy and hybridization-driven allopolyploidy. Therefore, as in other angiosperm lineages, the evolution of Brassicaceae genomes is characterized by cyclic rounds of WGDs followed by diploidization (Fig. 1.4b).

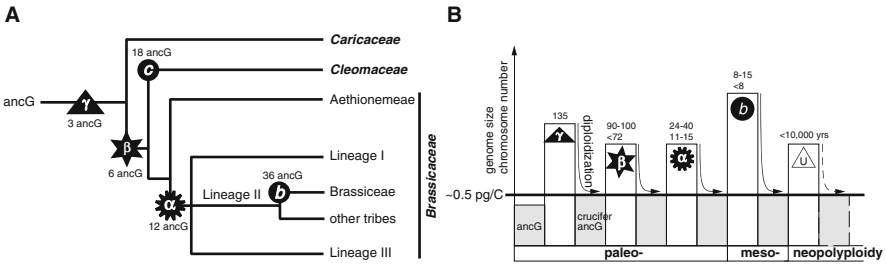

Fig. 1.4 **a** Discussed whole-genome duplication (WGD) events mapped onto the schematic phylogeny of the Brassicales. Besides γ and β events, lineage-specific tetraploidization α (core Brassicaceae) and whole-genome triplications *c* (Cleomaceae) and *b* (Brassiceae) are shown. The assumed number of ancestral genomes (ancG) multiplied by WGD events is also given. Modified after Beilstein et al. (2006), Lysak et al. (2006), Schranz and Mitchell-Olds (2006), Jaillon et al. (2007), Ming et al. (2008), and Tang et al. (2008). **b** Model of cyclic genome evolution in Brassicaceae. Multiple ancient and more recent WGDs (γ, β, α, and *b*; the U event refers to the origin of *Brassica* allopolyploids of the U's *triangle*) increasing genome size and chromosome number are counteracted by the genome diploidization, reducing genome size and chromosome number. For simplicity, the model assumes that polyploid genomes are being diploidized toward the reconstructed ancestral crucifer genome of ~0.5 pg/C (Lysak et al. 2009). The size of the ancestral angiosperm genome triplicated by the γ event was chosen arbitrarily. Rough time estimates of the WGDs are given in million years ago following Rana et al. (2004), Lysak et al. (2005), Henry et al. (2006), Franzke et al. (2008), and Ming et al. (2008)

Divergence time estimates (Fig. 1.3) are still controversial. The usage of K_s values, as presented by Schranz and Mitchell-Olds (2006) and Maere et al. (2005), might be more reliable since they do not make assumptions about molecular clocks. K_s values represent the synonymous substitution rate that can be inferred for protein-coding genes only. It is the calculation of a rate considering only synonymous mutations not resulting in amino acid changes. Schranz and Mitchell-Olds (2006) estimated the time of very early radiation of Brassicaceae as 34 mya. This was based on a genome-wide (referring to *A. thaliana*) estimated K_s average (K_s = 0.67) reflecting the most recent (α) WGD event and using *A. thaliana* as a reference (Simillion et al. 2002, Bowers et al. 2003, De Bodt et al. 2005). Genome-wide comparison of K_s values from Cleomaceae and Brassicaceae suggests that the corresponding mean K_s value is 0.82, which indicates 41 mya as a divergence time estimate for these two families. This range of divergence fits well the assumption that the split between Caricaceae and the core Brassicales had occurred 72 mya (Wikström et al. 2001). It is noteworthy that the divergence time estimates were obtained by different methods (calculating synonymous mutation rates and running simulation models, Koch et al. 2000; calculating genome-wide K_s values, Schranz and Mitchell-Olds 2006; or using relaxed molecular clock approaches, Franzke et al. 2009). Despite a relatively large variation, it is remarkable to see that the order of magnitude is similar. This is even the case for the divergence data provided by Franzke et al. (2009) showing much more recent estimates and which have been, however, re-calculated by the same authors providing much older estimates (Couvreur et al. 2009). A critical aspect is dating and calibration of the phylogenetic trees as the Brassicaceae taxa are poorly represented as macrofossils (Appel and Al-Shehbaz 2002). Some studies relied partly on macrofossils from the Miocene (Koch et al. 2000), whereas other relied on the Turonian fossil taxon *Dressiantha* (ca. 89.5 mya; Gandolfo et al. 1998) with its set of characters such as the presence of a gynophore, unequal petal size, or a bicarpellate gynoecium suggesting an affinity with the order Brassicales sensu Angiosperm Phylogeny Group (APGII 2003) (Magallon et al. 1999, Magallon and Sanderson 2001).

1.3.2 Genome Size Variation

Multiple whole-genome duplications and triplications have necessarily increased genome size (GS) of ancestral Brassicaceae species. Nevertheless this tendency is not apparent when GS of the extant crucifer taxa are surveyed. Despite the 16-fold variation in GS across the family, all crucifer species analyzed so far are classified as having very small ($\leq$ 1.4 pg) or small ($\leq$ 3.5 pg/C; Leitch et al. 1998) genomes, with a mean *C* value of 0.63 pg (Lysak et al. 2009). The smallest *C* values have been estimated for *A. thaliana* (0.16 pg/C; Camelineae) and some *Sphaerocardamum* species (0.15–0.16 pg/C; Halimolobeae). On the opposite pole, *Bunias orientalis* has the largest genome (2.43 pg/C; Buniadeae), followed by high GS values in some species of *Physaria* (2.23–2.34 pg/C; Physarieae) and *Matthiola* (2.11–2.29 pg/C; Anchonieae) (Lysak et al. 2009).

Two recent studies (Oyama et al. 2008, Lysak et al. 2009) attempted to disentangle the evolution of GS in Brassicaceae using available phylogenetic frameworks. Both studies revealed increases as well as decreases in GS across the analyzed species sets without apparent evolutionary directionality. Lysak et al. (2009) reconstructed a theoretical ancestral genome size (ancGS) as 0.5 pg/C and tested modes of GS evolution along phylogenetic lineages based on five gene markers. When comparing the *C* values, 50% of crucifer taxa analyzed showed a GS decrease and increase, respectively, and GS evolution shows no dominant tendency. Overall GS in the extant species as compared to the ancGS remained relatively stable across evolutionary time (Fig. 1.4b) and increases were generally moderate, with significant increases found only in the Anchonieae, Buniadeae, and Physarieae. Hence, despite dynamic processes having the potential to increase GS (e.g., transposable element amplification and WGDs), mechanisms eliminating accumulated DNA and/or suppressing DNA amplification must be active in Brassicaceae. Although possible mechanisms of GS downsizing were suggested [e.g., illegitimate recombination: Devos et al. (2002) or gross chromosome rearrangements: Lysak and Lexer (2006), Gaeta et al. (2007)], the modes of GS stasis in Brassicaceae are still poorly understood.

Particularly intriguing is genome evolution in crucifer species with large C values (2.0 pg/C), large chromosomes, and low, diploid-like chromosome number ($2n = 8$–14). This group of taxa thus far represents two *Bunias* ($2n = 14$), one *Physaria* ($2n = 8$), and three *Matthiola* ($2n = 14$) species (Lysak et al. 2009). To resolve this, it was proposed that the diploid-like number and large chromosome size may have arisen from ancient polyploidy events followed by diploidization involving chromosome number reduction without extensive DNA loss, or via retrotransposon amplification equally increasing size of all chromosomes as well as GS of the entire genome (Lysak et al. 2009). Although the polyploidization–diploidization scenario seems to be more conceivable, the issue cannot be resolved at present.

1.3.3 Chromosomes and Chromosome Number Variation

Mitotic chromosomes of crucifer species are generally very small, spanning only a few micrometers in size (e.g., 1.5–2.8 μm in *A. thaliana*, Koornneef et al. 2003, 2–5 μm in *Brassica*, Cheng et al. 1995; 1.2–5.8 μm in *Boechera*, Kantama et al. 2007). As noted before, larger chromosomes seem to be frequently observed in species with low chromosome numbers ($n = 4$–7) and large genome sizes (e.g., *Bunias*, *Matthiola*, *Physaria*; Manton 1932, Lysak et al. 2009), although large chromosomes were also reported in *Menonvillea* with $2n = 22$ (Manton 1932).

Chromosome morphology and structure of most of the crucifer species are closely linked to the discrete distribution of repetitive DNA elements along a longitudinal chromosome axis. The highest percentage of dispersed and tandem repeats including ribosomal DNA repeats constitutes distinct heterochromatic

arrays (Fig. 1.5). Euchromatic chromosome regions largely coincide with gene-rich sequences, though they are not completely depleted of dispersed repeats. Differential staining techniques reveal pericentromere regions as a prominent heterochromatic component of crucifer chromosomes. Both dispersed and satellite repeats compose interstitial and terminal heterochromatic knobs. Similar knobs known as nucleolar organizing regions (NORs) are formed by tandem arrays of 45S ribosomal DNA (rDNA). Distinct chromosome organization also determines the characteristic chromatin organization within interphase nuclei with distinct heterochromatic chromocenters comprising pericentromeres, knobs, and NORs, interspersed by euchromatic chromosome territories (Fransz et al. 2002, Koornneef et al. 2003, Pecinka et al. 2004).

Chromosome numbers, recently compiled by Warwick and Al-Shehbaz (2006), are known for ca. 70% of genera and 40% of crucifer species. Chromosome numbers in Brassicaceae vary over 32-fold, with the lowest chromosome number of $n = 4$ found only in *Physaria* (Physarieae) and *Stenopetalum* (Camelineae); five chromosome pairs ($n = 5$) were observed in *A. thaliana* (Camelineae), one *Matthiola* (Anchonieae), five *Stenopetalum* and 21 *Physaria* species (Warwick and Al-Shehbaz 2006). The highest chromosome numbers have been reported by Montgomery (1955), Easterly (1963), and Harriman (1965) in North American polyploid *Cardamine* (formerly *Dentaria*) species (*Cardamine angustata*, *Cardamine concatenata*, *Cardamine diphylla*, *Cardamine dissecta*, and *Cardamine maxima*). Among these polyploid species the highest counts were obtained in *Cardamine concatenata* (*Delphinula laciniata*; $2n = \pm 240$ by Montgomery 1955 and $2n = 256$ by Easterly 1963) and *C. diphylla* ($2n$=ca. 256, Harriman 1965). However, the counts have to be considered a priori as approximate and inaccurate due to the clumping of numerous very small chromosomes (Harriman 1965). This is illustrated by more than a three-fold variation in chromosome number ($2n = 74–256$) between counts in different root tips in *C. diphylla* (Harriman 1965). Future analyses of DAPI-stained meiotic chromosomes coupled with GISH and/or FISH localization of centromeric satellite repeats can elucidate chromosome numbers and the origin of the high polyploid *Cardamine* species.

Base chromosome numbers vary from $x = 4–17$ with more than one-third of the taxa having karyotypes based on $x = 8$ (Warwick and Al-Shehbaz 2006), implying that $x = 8$ is most likely an ancestral chromosome number of the whole family. Base chromosome numbers are practical in recognizing diploids ($2x$) from higher ploidy levels ($3x$, $4x$, etc.) within a given taxon. However, frequent auto- and allopolyploid events increasing the number of chromosome sets have been followed by species- and lineage-specific chromosome reshuffling. Chromosome fusions and fissions are causing an intra-generic numeric variation known as descending and ascending dysploidy. Therefore, several crucifer genera (e.g., *Brassica*, *Cochlearia*, *Diplotaxis*, *Erysimum*, and *Physaria*) are polybasic, i.e., characterized by multiple base chromosome numbers (Warwick and Al-Shehbaz 2006). This makes the base number concept impractical in some genera. Moreover, generic base numbers based on a lowest chromosome count available do not reflect the true nature of diploid-like

genomes which were often influenced by paleo- and mesopolyploid events followed by subsequent diploidization (see below).

1.3.4 Hybridization and Polyploidy

Like in other plant families, polyploidy, hybridization and introgression have significantly shaped genome evolution in many crucifer species groups and genera. Two key review papers on polyploidy in Brassicaceae are available: Marhold and Lihová (2006) provided a comprehensive review on polyploid evolution in the family, and a chromosome number database was compiled by Warwick and Al-Shehbaz (2006). The importance of polyploid speciation is clearly demonstrated in some Brassiceae genera comprising exclusively polyploid species such as *Crambe* ($2n = 30$, 60, and 120), *Moricandia* ($2n = 22$, 28, 56, and 84), and *Vella* ($2n = 34$, 68, and 102) or in genera with polyploid taxa prevailing over diploids (e.g., *Aethionema*, *Braya*, *Draba*). The most detailed polyploidy pattern distribution has been elaborated for the genus *Draba* comprising more than 350 species and therefore representing one of the largest genera of Brassicaceae (Jordon-Thaden and Koch 2008). Here, polyploidization and speciation rates are exceeding by far those described for all other plants (Jordon-Thaden and Koch 2008). Numerous cases of inter-species hybridization and polyploid formation reported in Brassicaceae are evolutionary young events (ca. 10^2–10^5 years old) which should be distinguished from much older paleo- and mesopolyploid whole-genome duplications. In crucifers, new polyploid species are generated by autopolyploidy or through inter-species hybridization. Hybrid speciation includes either hybrids between genomes of different chromosome numbers, often undergoing subsequent duplication (allopolyploidy) or (perhaps rarer) hybrids between species with the same chromosome numbers (homoploid hybrid speciation or recombinational speciation) (see Mallet 2007, Soltis et al. 2007, Hegarty and Hiscock 2008 for reviews).

Drawing a distinct line between autopolyploidy and allopolyploidy is not always straightforward and in some cases not feasible due to a problematic definition of a species (cf. Rieseberg and Willis 2007, Mallet 2007). Numerous reports on even-numbered intraspecific cytotypes (cf. Warwick and Al-Shehbaz 2006; e.g., *Calepina irregularis*: $2n{=}2x$, $4x{=}16$, 32; *Isatis tinctoria*: $2n{=}2x$, $4x{=}14$, 28) are suggesting that autopolyploidy is common in Brassicaceae. This is consistent with the present view on autopolyploidy as a significant, though underestimated mode of polyploid speciation in plants (Soltis et al. 2007). Taxonomic treatment of autopolyploid derivates and their respective progenitors is problematic. Most frequently diploid and autopolyploid populations possess indistinguishable phenotypes and are treated as cytotypes without a taxonomic rank (e.g., diploid and autopolyploid cytotypes within four species of the *Cardamine digitata* aggregate, Jørgensen et al. 2008); less common is their circumscription as subspecies (autotetraploid *Biscutella laevigata* subsp. *laevigata*, $2n{=}4x{=}36$, Tremetsberger et al. 2002, autotetraploid *Cardamine amara* subsp. *austriaca*, $2n{=}4x{=}32$, Marhold 1999) or separate species (*Cochlearia officinalis*, $2n{=}4x{=}24$ and *C. anglica*, $2n{=}8x{=}42$; Koch et al. 1998a,

1999b; *Cardamine matthioli*, 2*n*=2*x*=16 and *C. majovskii*, 2*n*=4*x*=32; Lihová and Marhold 2003). The most cited example of autopolyploid speciation in Brassicaceae is that of the tetraploid *B. laevigata* described by Manton (1937). *B. laevigata* is the most common *Biscutella* species comprising diploid (2*n*=2*x*=18) and tetraploid (2*n*=4*x*=36) cytotypes. The diploid cytotype is treated as several subspecies, whereas the tetraploid populations confined to higher altitudes of the Alps were circumscribed as subsp. *laevigata* (Tremetsberger et al. 2002, Parisod and Besnard 2007). As the tetraploid populations show a tetrasomic inheritance and morphologically resemble different diploid subspecies, it was concluded that the tetraploid taxon originated via a polytopic autopolyploidy (Tremetsberger et al. 2002). The autopolyploid origin is also underlined by the occurrence of tetravalents and trivalents during meiosis I (Manton 1937). However, the multivalent formation should not be taken as a conclusive evidence of an autopolyploid origin as some autopolyploids can exhibit regular bivalent pairing and still display polysomic inheritance (Santos et al. 2003, Soltis et al. 2007). Analysis of meiotic pairing in established and newly generated *A. thaliana* autotetraploid (2*n*=4*x*=20) lines, revealed concomitant decrease and increase in multivalents and bivalents, respectively, suggesting a partial diploidization of meiosis (Santos et al. 2003). In presumably autotetraploid *Polyctenium fremontii* (2*n*=4*x*=28, Boechereae), although tetravalents are frequently observed in pachytene, regular bivalents prevail in metaphase I (T. Mandáková and M.A. Lysak, unpublished data). A regular bivalent pairing throughout the first meiotic division has been found in two apparently autotetraploid (2*n*=4*x*=28) species of the Calepineae, *C. irregularis*, and *Goldbachia laevigata* (Mandáková and Lysak 2008). As cross-species chromosome painting allows the identification of individual chromosomes during meiosis, pairing of homologous and homeologous chromosomes in autopolyploid plants can be analyzed with a high degree of precision.

Inter-species hybridization occurs frequently in Brassicaceae, particularly in some genera such as *Arabis*, *Boechera*, *Cardamine*, or *Rorippa*. Hybrids can be sterile, partially, or fully fertile and then having the capacity of backcrossing to the respective parental species. For instance, long-distance dispersal followed by hybridization has been suggested for the origin of the polyploid Australian and New Zealand *Lepidium* taxa. Phylogenetic analysis of noncoding chloroplast and nuclear (ITS) DNA regions showed that Australian/New Zealand taxa have most likely an allopolyploid and bicontinental origin, resulting from crosses between African and Californian *Lepidium* species (Mummenhoff et al. 2004). Other well-documented examples where interspecific hybrids have become newly established allopolyploid species include the composite genomes of *Diplotaxis muralis* (2*n*=42; parental species *Diplotaxis viminea*, 2*n*=20 and *Diplotaxis tenuifolia*, 2*n*=22; Eschmann-Grupe et al. 2004), three *Brassica* allotetraploids of the U's triangle (e.g., 2*n*=38 in *B. napus*, parental subsp. *B. oleracea*, 2*n*=18, and *Brassica rapa*, 2*n*=20), *Arabidopsis kamchatica* (2*n*=32; parental subsp. *A. lyrata* and *A. halleri* subsp. *gemmifera*, both 2*n*=16; Shimizu et al. 2005, Koch and Matschinger 2007, Schmickl et al. 2008), or *A. suecica* (2*n*=26; parental subsp. *A. arenosa*, 2*n*=16, 32 and *A. thaliana*, 2*n*=10; Mummenhoff and Hurka 1994). In some hybrids, the

genome incompatibility has been overcome by whole-genome duplication (autoallopolyploidy), either directly or via a triploid bridge and unreduced gametes. To such allopolyploid crucifer species belong *Draba ladina* (2*n*=32; parental subsp. *Draba aizoides* and *Draba tomentosa*, both 2*n*=16; Widmer and Baltisberger 1999) or *Cardamine schulzii* (2*n*=6*x*=48), an autopolyploid derivate of the interspecific hybrid (*C.* × *insueta*, 2*n*=3*x*=24) between *C. amara* and *Cardamine rivularis* (both 2*n*=16) (Urbanska et al. 1997). In the genus *Boechera*, the potential sterility of the hybridogenous *B.* × *divaricarpa* (*Boechera holboellii* × *Boechera stricta*) and its negative evolutionary consequences have been overcome by the apomictic reproduction (Dobes et al. 2004a, b, Schranz et al. 2005). In European *Microthlaspi perfoliatum* with 2*n*=14, 28, 42 cytotypes the situation is less clear and one of the putative parental taxa of the hexaploid cytotype most likely went extinct (Koch et al. 1998b, Koch and Hurka 1999, Koch and Bernhardt 2004).

Genomic in situ hybridization (GISH) refers to the specific identification of parental genomes in interspecific hybrids and allopolyploid species by simultaneous or subsequent hybridization of fluorescently labeled genomic DNA (gDNA) of the presumed parents to chromosomes of the composite genome. In Brassicaceae, most GISH studies were initially concentrated on the natural and synthetic allotetraploid *Brassica* species (*Brassica carinata*, *Brassica juncea*, *B. napus*) and artificial intergeneric hybrids between *B. napus* and other crucifer species mainly from the Brassiceae (Chèvre et al. 2007, reviewed by Lysak and Lexer 2006, Snowdon 2007). GISH data corroborated the phylogenetic relationships among the three "diploid" donor *Brassica* species with the two lineages separating *Brassica nigra* (BB genome, 2*n*=16) from *B. oleracea* (CC, 2*n*=18) and *B. rapa* (AA, 2*n*=20) (Warwick and Sauder 2005 and references therein). GISH discriminated between A and C genomes in *B. carinata* (BBCC, 2*n*=34) and between A and B genomes in *B. juncea* (AABB, 2*n*=36) (Snowdon et al. 1997, Maluszynska and Hasterok 2005), whereas A and C genomes showing a high level of sequence homeology have not been discerned in *B. napus* (AACC, 2*n*=38) (Snowdon et al. 1997). This was not confirmed by Howell et al. (2008) who showed that the A and C genomes can be distinguished clearly using gDNA of *B. oleracea* as the probe and gDNA of *B. rapa* DNA as a block. Despite a very close phylogenetic relationship between *Arabidopsis thaliana* (TT, 2*n*=10) and *Arabidopsis arenosa* (AA, 2*n*=16; AAAA, 2*n*=32), GISH with increased amounts of gDNAs was successfully applied to identify both parental genomes in the natural and synthetic *Arabidopsis suecica* (AATT, 2*n*=26) as well as in an artificial hybrid between *Arabidopsis thaliana* and *A. suecica* (Ali et al. 2004, Lysak and Lexer 2006). Recently, GISH was used to elucidate the genome composition of sexual (2*n*=14) and apomictic (2*n*=15) genotypes from the *Boechera holboellii* complex (Kantama et al. 2007). This species complex, including *B. holboellii*, *B. stricta*, and their presumed hybrid *B.* × *divaricarpa*, exhibits extensive karyological variation due to recurrent hybridization, introgression, and apomixis. Two-color GISH analysis using *B. holboellii* and *B. stricta* gDNAs revealed that the analyzed apomicts represent interspecific hybrids with different contribution of *B. holboellii*- and *B. stricta*-derived chromosomes (Kantama et al. 2007). As in most

studies carried out for Brassicaceae species so far, GISH in *Boechera* was principally based on genome-specific pericentromeric tandem repeats (Kantama et al. 2007, see also Ali et al. 2004, Lysak and Lexer 2006).

These examples of GISH studies illustrate an untapped capacity of the technique to reveal the origin of hybridogenous taxa within taxonomically complicated crucifer groups (e.g., *Aethionema*, *Cardamine*, *Draba*, *Heliophila*; also see Marhold and Lihová 2006).

1.3.5 Genome and Chromosome Collinearity

Based on DNA markers, molecular phylogenetics is steadily improving our understanding of taxon-to-taxon relationships within Brassicaceae. Nevertheless, inter-species and infrafamiliar relationships may also be inferred from genome-wide comparisons. Such comparisons can be carried out through comparative genetic mapping, comparative cytogenetic analysis, or by comparing whole-genome sequences. As comparative genetic mapping in Brassicaceae is extensively covered by other Chapters 5 and 6 and recent reviews (e.g., Koch and Kiefer 2005, Lysak and Lexer 2006, Schranz et al. 2006, Snowdon 2007), only a brief account of key findings directly linked to genome and karyotype evolution of Brassicaceae is outlined herein.

Arabidopsis sequence data along with the wealth of genetic markers have been crucial for *Arabidopsis* becoming a reference genome in comparative genetic mapping across Brassicaceae.

The first wave of crucifer comparative genetics was marked by analyzing the extent of cross-species genome collinearity between the diploid *Brassica* species and between *Arabidopsis* and *Brassica* species, respectively (Lagercrantz and Lydiate 1996, Lagercrantz 1998). However, the budding field of whole-genome comparisons in Brassicaceae has been fostered significantly by genetic maps comparing *Arabidopsis* with three $n=8$ Camelineae taxa, *A. lyrata* subsp. *petraea* (Kuittinen et al. 2004), *A. lyrata* subsp. *lyrata* (Yogeeswaran et al. 2005) and *C. rubella* (Boivin et al. 2004). Despite some discrepancies among the three Camelineae karyotypes, particularly in the number of inferred inversion events due to the different marker density, all three maps were largely congruent and showed a strikingly high extent of large-scale collinearity with the five *A. thaliana* linkage groups (Boivin et al. 2004, Kuittinen et al. 2004, Koch and Kiefer 2005, Yogeeswaran et al. 2005, Lysak and Lexer 2006).

1.3.6 Revealing Chromosome Homeology Through Comparative Chromosome Painting

Chromosome painting points to the identification of large chromosome regions or whole chromosomes using chromosome-specific DNA probes. Fluorescently