Chapter 5
Chasing Ghosts: Comparative Mapping in the Brassicaceae

Isobel Parkin

Abstract The study of plant genome organization has benefited greatly from the application of comparative genetic mapping, which allows both the elucidation of chromosomal rearrangements resulting from speciation and the ability to transfer information and resources between species. A significant focus of comparative mapping in the Brassicaceae has been within the agronomically important species of the *Brassica* genera and between the *Brassica* crops and their well-characterized relative *Arabidopsis thaliana*. These studies have demonstrated the ghostly remnants of an hexaploid ancestor in the evolutionary past of the *Brassica* diploids that explain the observed levels of gene duplication within the genomes. Further, comparative mapping with *A. thaliana* has uncovered a segmental architecture of conserved ancestral blocks which can be replicated and rearranged to reflect the current genomes of all members of the Brassicaceae studied to date. The correspondence between the *A. thaliana* and *Brassica* genomic regions is being exploited to fine map, identify, and clone genes for economically valuable traits.

Keywords Homology · Collinearity · Conserved genome blocks · Polyploidy · Chromosomal rearrangements

Contents

I. Parkin (✉)
Agriculture and Agri-Food Canada, Saskatoon Research Centre, Saskatoon, SK, Canada S7N 0X2
e-mail: isobel.parkin@agr.gc.ca

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_5,

5.1 Introduction

Early efforts to understand plant genome organization and evolution relied upon coarse imaging of whole chromosome structures through microscopy. By interpreting the pairing structures formed during meiosis it was possible to infer how closely related chromosomes, and hence, genomes and species were. The advent of molecular markers and in particular in the 1980s the ability to visualize restriction fragment length polymorphisms (RFLP) as a tool in genetic linkage analysis enabled more comprehensive plant genome analysis and heralded the dawn of comparative mapping (reviewed in Gale and Devos (1998)).

The premise was simple: that common sets of markers could be used to identify related segments of DNA between not only species of the same tribe but also across wide taxonomic distances. Such analyses could even transcend the millions of years represented by the monocot–dicot species divide (Paterson et al. 2004, Tang et al. 2008). These data showed extensive conservation of gene content between species of the same tribe and it was demonstrated that taxonomically disparate species could in some instances be separated by a remarkably limited number of major or large effect chromosomal rearrangements. For example, in the Solanaceae, the tomato and potato genomes were differentiated by a mere five inversion events (Tanksley et al. 1992). The most comprehensive analysis of a single family has been completed for the Poaceae (reviewed in Devos and Gale (2000)), where the concept of ancestral blocks was first proposed (Moore et al. 1995).

With a relatively limited number of markers it was shown that the genome of the monocot model rice could be broken down into 19 conserved blocks of collinearity which upon rearrangement formed the genomes of a diversity of cereal species, despite a wide range in base chromosome number, genomic DNA content, and estimated divergence times of up to 60 million years. The addition of further markers and comparisons between additional species have more accurately defined the ancestral blocks, increasing the number to 30 and culminating in the "circle of cereals," an unified comparative map of the grasses, where the rice genetic linkage map is drawn at the origin of a set of concentric circles, each circle representing an additional cereal genome. This novel representation allows the relatively simple identification of related genome segments across species (Devos 2005).

In the Brassicaceae a number of genetic linkage maps have been generated over the past 20 years, largely focusing on the agronomically important members of U's triangle and utilizing different sets of markers that prevented cross-species comparison (Slocum et al. 1990, Landry et al. 1991, Kianian and Quiros 1992, Ferreira

Table 5.1 Description of selected published genetic linkage maps which have contributed to our understanding of genome evolution and organization within the *Brassica* genus of lineage II

Target species	Genomes under comparison[a]	References	Number of conserved loci (markers)[b]	Length of genetic map for target species (cM)
B. rapa (A)	$A–A^{Bna}$	Suwabe et al. (2008)	(44)	743
B. oleracea (C)	$C–C^{Bna}$	Bohuon et al. (1996)	129	875
B. oleracea (C)	C–At	Lan and Paterson (2000)	186	863.6
B. oleracea (C)	C–At	Lukens et al. (2003)	(131)	875[c]
B. nigra (B)	$B–A^{Bna}–C^{Bna}$	Lagercrantz and Lydiate (1996)	(158)	778
B. nigra (B)	B–At	Lagercrantz (1998)	284–160	751
B. napus (AC)	$A^{Bna}–C^{Bna}$	Parkin et al. (2003)	(162)	1,698.5
B. napus (AC)	$A^{Bna}–C^{Bna}$–At	Parkin et al. (2005)	1,232–550 (368)	1,968
B. juncea (AB)	$A^{Bj}–B^{Bj}–A^{Bna}–B^{Bni}–C^{Bna}$–At	Panjabi et al. (2008)	533	1,992.2

[a]The constituent genomes are indicated: A, B, or C for the diploids *B. rapa*, *B. oleracea*, and *B. nigra*, respectively; superscript *Bna*, *Bj*, or *Bni* indicate the respective diploid genome within the allopolyploid nucleus of *B. napus*, *B. juncea*, or the diploid *B. nigra*.
[b]The number of markers used to detect comparative loci is indicated in brackets. However, due to duplication within the genomes this will not necessarily reflect the number of conserved loci used in the comparison, which where available is preferentially indicated.
[c]The *B. oleracea* map used in Lukens et al. (2003) was taken from Bohuon et al. (1996).

et al. 1994, Uzunova et al. 1995, Piquemal et al. 2005). However, in the late 1990s the application of a common set of markers across the three diploid species and the amphidiploid genome derivatives allowed comparisons to be made within the *Brassica* genus of lineage II of the Brassicaceae (Table 5.1). More recently the sequencing of previously mapped RFLP probes and the use of sequence-based markers have identified related genomic regions of the dicot model *Arabidopsis thaliana* (lineage I) and defined a set of ancestral blocks which allow similar comparisons as those previously made in the Poaceae (Parkin et al. 2005, Panjabi et al. 2008). The knowledge gained from these comparative mapping studies in the Brassicaceae will be described, focusing largely on the *Brassica* genus, and more recent and future developments relevant to interpreting *Brassica* genome organization will be introduced.

5.2 Common Terms Used in Comparative Mapping Studies

The wide adoption of comparative mapping as a tool to study plant evolution has inevitably led to the creation of a new vocabulary to describe common themes. In some instances established genetics nomenclature has been appropriated and imbued with subtle differences of meaning, which can lead to confusion in the literature. The definition of some of the more commonly used terms is provided below.

Synteny: This describes the physical co-localization of genetic loci on the same chromosome within an individual or species and is often erroneously used in place of collinearity; syntenous loci although always physically linked are not necessarily genetically linked or arranged in a predictable pattern.

Collinearity: Strictly speaking this refers to multiple points found in a linear order; in comparative mapping "conserved linkage" or collinearity refers to regions of conserved marker content and order on two (or more) separate linkage groups (or chromosomes).

Homology: In the context of comparative mapping it is generally used to refer to chromosomal regions but sometimes individual genetic loci and indicates the shared ancestry of these homologous elements.

Homoeology: Refers to chromosomes, regions, or loci inherited from divergent but homologous genomes within an allopolyploid nucleus.

Polyploidy: Mode of evolution which involves the doubling of genome complements either through whole genome duplication (autopolyploidy) or through hybridization of two or more related but distinct genomes (allopolyploidy).

Orthologues and paralogues: The identification of orthologues, those sequences related by speciation, and paralogues, those sequences which originate through segmental or gene duplication, is contentious due to the dynamic nature of genome evolution. For example, the paralogue maybe favored and the "true" orthologue of a gene is lost over time due to adaptive pressures. To limit such confusion the more general term of homology will be largely used throughout the text.

5.3 The Basics of Comparative Mapping

Comparative mapping is a powerful tool which not only allows the study of genome evolution but also can be exploited to transfer knowledge and resources from model plant species to improve traits in related crops. Much of the comparative mapping work to date has utilized RFLP markers. Restriction fragment length polymorphism (RFLP) loci are revealed as differences in genomic fragment lengths resulting from sequence variation at enzyme restriction sites; RFLP probes are derived from labeled cDNA or genomic DNA and through hybridization identify homologous sequences within genomes, thus allowing the length polymorphisms to be visualized while simultaneously identifying conserved sequences across species. The relationship between genetic linkage maps is identified through the use of common sets of RFLP markers and the similarity of the species is assessed through the extent of conservation of marker content and order or collinearity (Fig. 5.1). The level of collinearity varies with the rates and modes of evolution which is unique to each plant lineage, ranging from almost complete alignment of linkage groups to the identification of many small blocks of similarity between genomes (Fig. 5.1). The breaks in the collinearity are indicative of chromosomal rearrangements, such as duplications, translocations, inversions, or transpositions, and detail the evolutionary history of each species. In any comparative mapping study, caution should

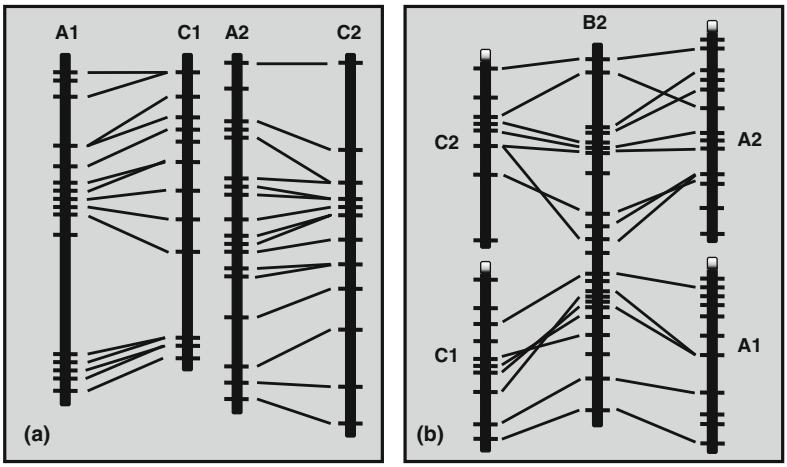

Fig. 5.1 Utilizing molecular markers to identify collinearity and uncover genome rearrangements. Comparing closely related species will identify limited rearrangements and greater stretches of conserved marker content and order, in (**a**) the *B. napus* A genome linkage groups A1 and A2 and their C genome homoeologues C1 and C2 are shown, these groups are collinear along their entire mapped lengths. However, when considering more distantly related species, for example, in (**b**) the B genome linkage group B2 (G5) is compared to the same *B. napus* linkage groups as in (**a**), the impacts of both genome duplication, B2 (G5) is collinear with segments of both A1 and C1, and chromosomal rearrangements such as translocations, chromosome fissions/fusions, and inversions are observed

be exercised due to the limitations imposed by genetic linkage analysis, in that only polymorphic regions of the genome can be observed and physical restriction of chromosome pairing can create clusters of coincident markers impeding the elucidation of marker order and hence collinearity.

Some of these limitations have been assuaged with the availability of fully sequenced genomes as points of reference in comparative mapping studies. For example, in the Brassicaceae, probes used for genetic mapping can be physically positioned on the *A. thaliana* genomic sequence based on conservation at the DNA sequence level. However, the question becomes what level of identity between two sequences indicates an evolutionary determined relationship and what is the impact of genome duplication, how similar are orthologues and paralogues found within the same genome to the ancestral homologue? By comparing the sequences of marker probes previously mapped in *B. oleracea* with genomic sequence from the crucifer model *A. thaliana,* Lukens et al. (2003) was able to define criteria by which to accept or reject the identification of orthologues (or the primary homologue) between the two species. These criteria were based on the observed distribution of sequence similarity scores (BLASTN) and the a priori knowledge of the expected nucleotide differences between paralogous (secondary homologues) sequences within the model genome itself (Lynch and Conery 2000), defining a significance of sequence similarity cut-off above which conserved sequences are

expected to represent orthologues. This assumes that a query sequence from one genome is orthologous only to the other genome's sequence to which it is most similar and although not infallible, since strongly conserved motifs from replicated gene families can confound such analyses, such criteria provide a foundation for delineating conserved regions between genomes.

5.4 The Contribution of Polyploidy (Inter-specific Hybridization) to *Brassica* Genome Evolution

Polyploidy has played a significant role in the evolution of the Brassicaceae as it has done in many plant lineages. In the 1930s the eponymous U's triangle, as described in Chapter 1, was determined from cytological analyses of forced inter-specific hybrids (U 1935). Genetic linkage analysis has since confirmed the relationship between the six domesticated *Brassica* species, demonstrating at the molecular level that the three diploid species, *B. rapa* (A genome), *B. oleracea* (C), and *B. nigra* (B), had formed the three amphidiploid species, *B. napus* (AC), *B. juncea* (AB), and *B. carinata* (BC), through each possible pair-wise combination (Parkin et al. 1995, Bohuon et al. 1996, Panjabi et al. 2008, Suwabe et al. 2008).

The generation of new polyploids requires rapid diploidization within the nucleus to ensure stable pairing and inheritance; this can be achieved through significant chromosome rearrangement accentuated by sequence elimination or through genetic suppression of pairing between non-homologous chromosomes (or homoeologues) (Jenczewski and Alix 2004). Genetic linkage maps have been generated for *B. napus* and *B. juncea* as a result of normal diploid pairing between chromosomes of established allopolyploid lines and their modern day diploid relatives (Parkin et al. 1995, Axelsson et al. 2000). This has allowed the two diploid genomes to be identified within the allopolyploid nucleus and indicated that no major chromosomal rearrangements have occurred since the fusion of the *Brassica* A genome with either the B or the C genomes. This suggests that similar to wheat the *Brassica* allopolyploids thrived through the evolution of a heritable mechanism that suppresses illegitimate recombination events. However, although this is thought to be true for *B. napus*, where relatively high levels of homoeologous (or non-homologous) pairing have been observed in newly resynthesized lines, no such pairing was observed for resynthesized *B. juncea*. It appears that the divergence of the B genome from that of the A and C through both chromosomal rearrangements (see below) and genetic drift has been sufficient to limit pairing across these species, although it cannot be ruled out that the B genome retains a strong suppressor of illegitimate pairing. The linkage maps of the allopolyploids and those of the modern progenitor species represented the first comparative mapping data for the Brassicaceae.

To enable comparative mapping between species which have evolved through genome duplication ideally it is necessary to compare only homologous regions but with increased levels of duplication this becomes progressively more difficult. The first attempts to compare the genome of *B. oleracea* with *B. napus* led to the erroneous conclusion that there had been significant rearrangement of the C genome within the *B. napus* nucleus upon fusion with the A genome progenitor (Cheung et al. 1997). This confusion was the result of comparing the diploid genome with

both the A and the C genomes within the allopolyploid species and thus identifying rearrangements which separated the two diploids. Such errors can be limited either through marker saturation or by inferring evolutionary relationships from evidence of pairing between similar chromosomes.

The homoeologous pairing events observed between the more closely related A and C genomes, identified through the utilization of a newly resynthesized *B. napus* line in mapping studies, have been exploited to identify the regions of primary homology between the two genomes (Parkin et al. 2003). These data were corroborated by the presence of extensive collinearity which determined that a minimum number of 16 chromosomal rearrangements were necessary to differentiate the A and C genomes (Fig. 5.2). The *Brassica* B genome has been shown to be phylogenetically distinct from the A and C diploids, which is reflected at both the sequence level (Sabhyata et al. 1996, Sharpe and Parkin, unpublished data) and in the absence of homoeologous pairing between the A and the B genomes (Axelsson et al. 2000). In contrast, although there remains some ambiguity in identifying the homologous regions across the three *Brassica* diploid genomes (A, B and C), comparative mapping has suggested a similar number of chromosomal rearrangements separate all three species (Fig. 5.2) (Lagercrantz and Lydiate 1996, Parkin et al. 2003, Panjabi et al. 2008). Also apparent from these initial studies of species from U's triangle was the high level of intra- and inter-chromosomal duplication observed within the *Brassica* diploid genomes suggesting the lineage had evolved from an ancient polyploid. This theory was cemented by subsequent comparisons to the genome of the

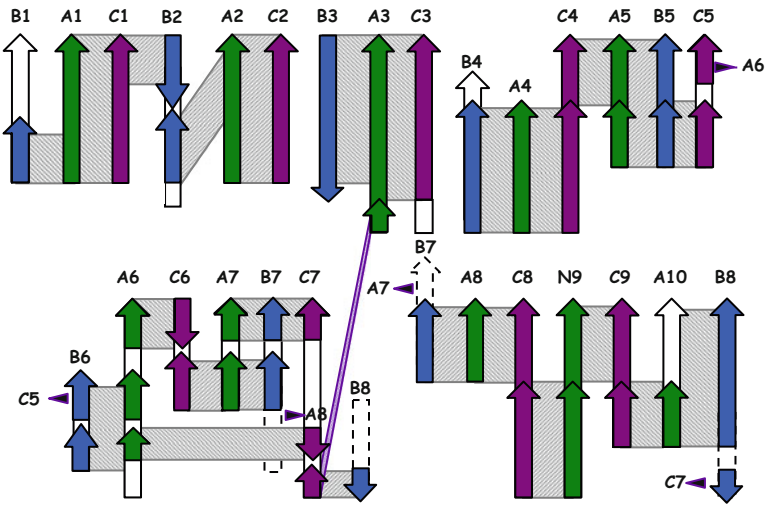

Fig. 5.2 Comparative mapping across the *Brassica* A (*green*), B (*blue*) and C (*purple*) diploid genomes. A schematic representation of the conserved genomic regions across the three genomes, the collinear segments are indicated by the *grey shaded regions* between linkage groups. The orientation of the linkage groups is indicated by the *arrows* (*pointing to top*) and is according to the published maps of Parkin et al. (2005) and Lagercrantz and Lydiate (1995). The B genome linkage groups are named according to the recent work of Panjabi et al. (2008)

plant model *A. thaliana*, a member of the Brassicaceae believed to have diverged from the *Brassica* species 14–24 million years ago (see Section 5.6 below) (Koch et al. 2000).

5.5 The Ghost of an Ancestral Hexaploid Genome

The use of RFLP markers to generate the first *Brassica* linkage maps was instrumental in uncovering evidence of whole genome duplication events within the diploid genomes. Up to 73% of RFLP probes have been shown to identify two or more loci within the diploid *Brassica* genomes (Lagercrantz and Lydiate 1996, Parkin et al. 2003). Further the duplicate loci were not randomly distributed across the genome but were found in collinear blocks of genetically linked loci. A recurring pattern was beginning to emerge with significant numbers of these conserved blocks being observed three times within the different genomes (Fig. 5.3). It should be noted that not all RFLP probes will reveal three loci in each diploid genome. Detection of all homologous regions is limited by not only the available polymorphism but also the on-going evolution of each conserved segment. Town et al. (2006) sequenced the three homologous regions from each of the triplicated copies of *B. oleracea* equivalent to two regions of the *A. thaliana* genome. This analysis provided the first comprehensive molecular study uncovering the impact of evolution upon the duplicated regions in *Brassica* genomes. It was demonstrated that each copy will be under local adaptive pressures that lead to sequence divergence and in some instances complete or partial loss of genomic sequences

Although the cumulative evidence from the *Brassica* diploids was highly suggestive, it was not until the first comparisons with the genome of the *A. thaliana* model plant were made that a hexaploid ancestor was proposed as the progenitor of the *Brassica* lineage (Lagercrantz 1998), an hypothesis which is now widely accepted.

5.6 *A. thaliana*, a Model Genome for the Brassicaceae

The benefactor of many decades of research, *A. thaliana*, emerged as the de facto model for plant species, which resulted in the development of extensive genetic and genomics resources including the first fully sequenced plant genome (Arabidopsis Genome Initiative 2000). Fortuitously for *Brassica* researchers, *A. thaliana* is closely related at the sequence level (~86% within coding regions) to modern day brassicas (Parkin et al. 2005); by sequencing RFLP probes previously mapped in *Brassica* species it was possible to identify through *in silico* analysis the most similar sequence(s) within the *A. thaliana* genome. Lukens et al. (2003) was the first study to use this approach and identified 34 regions of the *A. thaliana* genome with significant collinearity to almost 30% of the genetic map of *B. oleracea*. The increased density of markers, or possible points of comparison, developed for *B. napus* identified 21 blocks of genetically linked markers in the A and C genome of *B. napus* that were also found to be physically linked on the genome of *A. thaliana*

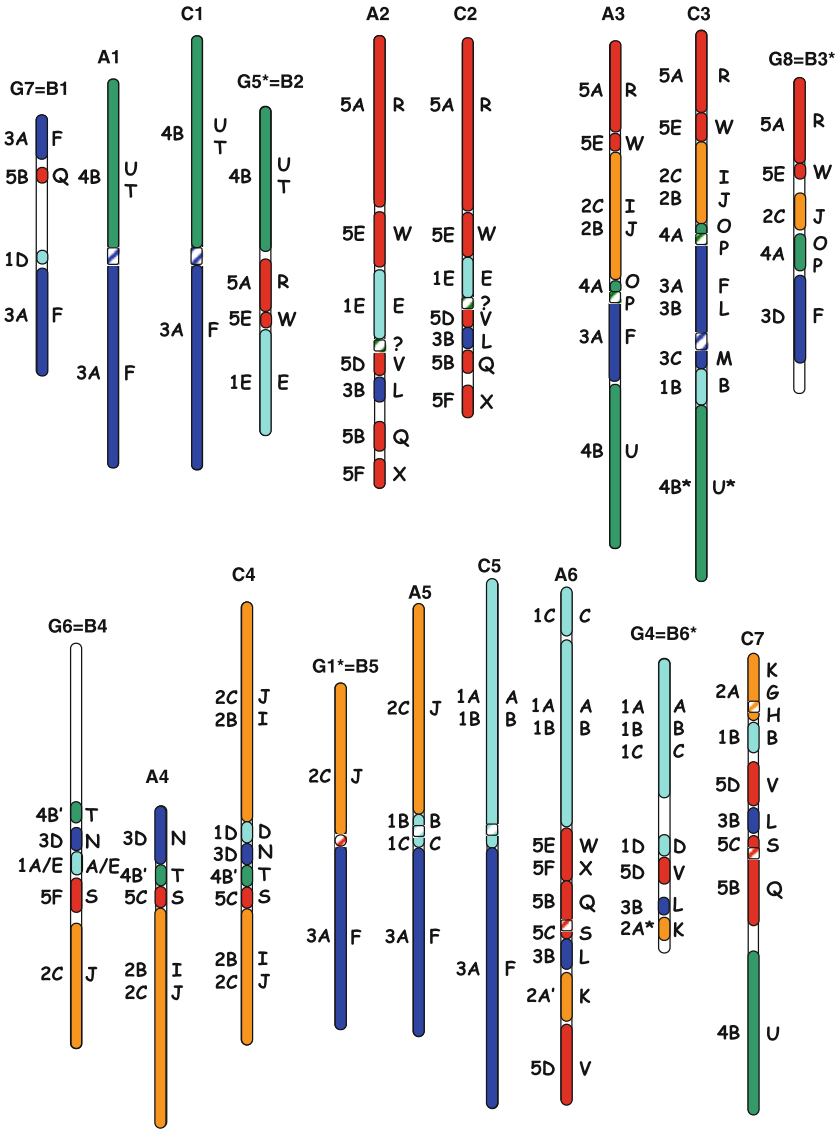

Fig. 5.3 How the ancestral blocks are arranged to construct the linkage groups of the A, B, and C genomes of *Brassica* species. The original designation of the blocks based on their collinearity with *A. thaliana* is shown to the *left* of the linkage groups (Parkin et al. 2005); the suggested block nomenclature according to collinearity with the lineage I ancestral karyotype is shown to the *right* (Schranz et al. 2006). The colors indicate collinearity with different *A. thaliana* chromosomes: *light blue* – chromosome 1; *orange* – chromosome 2; *dark blue* – chromosome 3; *green* – chromosome 4; and *red* – chromosome 5. The patterned blocks indicate the tentative positions of centromeres. The A and C linkage groups are based on the maps of Parkin et al. (2005) and the B genome is a derived composite based on information presented in Largercrantz and Lydiate (1995) and Panjabi et al. (2008), inverted linkage groups compared to these maps are indicated by an *asterisk*

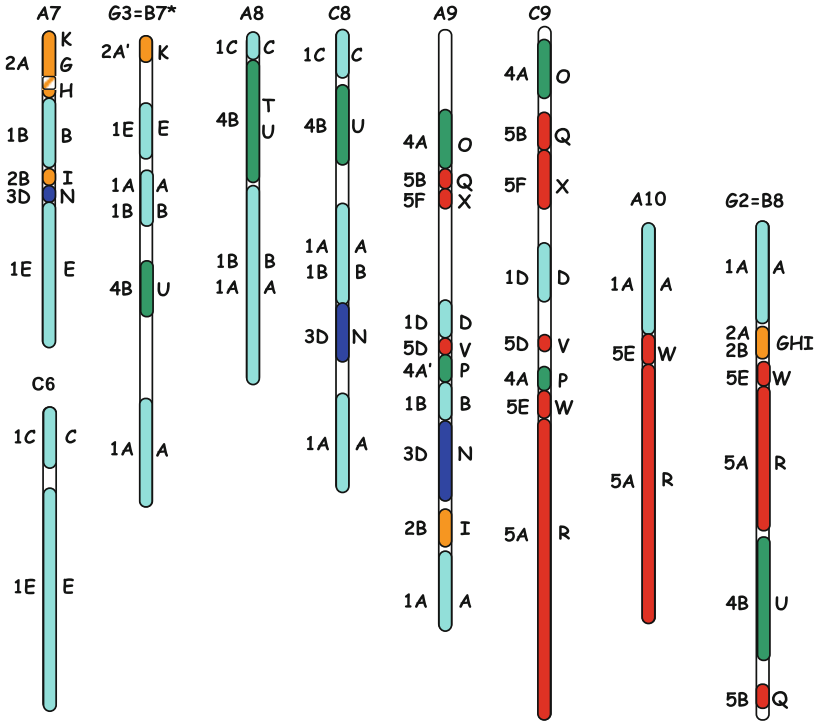

Fig. 5.3 (continued)

(Fig. 5.3) (Parkin et al. 2005). These conserved genomic units or ancestral blocks defined a framework that could be replicated and rearranged to represent over 80% of the *B. napus* genome. The mapping also corroborated the underlying triplicated nature of the *Brassica* genomes, with over 85% of the *B. napus* linkage map found in six copies. The fact that each region of the *A. thaliana* genome was found in six copies within *B. napus* indicated that the ancient duplication events which have taken place in the evolution of the *A. thaliana* genome (Blanc et al. 2003, Bowers et al. 2003, Henry et al. 2006) predate the triplication that has occurred in the *Brassica* lineage.

5.6.1 Across the A, B, and C Genomes

The comparison between the A and the C genomes has since been extended to the B genome to uncover a similar segmental arrangement of blocks (Panjabi et al. 2008). These data not only add to a previous study comparing *A. thaliana* with the B genome (Lagercrantz 1998) but also suggest the earlier work was flawed by the use of markers which later proved to be duplicated in the *A. thaliana* genome. There are still gaps in each of the diploid genomes where the relationship with *A. thaliana* is

tenuous due to a dearth of markers; however, the use of the block-based architecture across all three genomes presents an opportunity to uncover conserved arrangements of blocks common to the *Brassica* lineage, which may provide insights into the organization of the *Brassica* progenitor. For example, the inversion of block 5E (W) to lie next to block 5A (R) (present on A2/C2, A3/C3, C9/A10, and B3/B8: Fig. 5.3) is common to all three genomes. Interestingly this arrangement is also conserved in a number of $x = 7$ taxa from lineage I but is not present in the $x = 8$ taxa, the chromosomal organization of which had been proposed as an ancestral karyotype for both lineage I and II species (Mandakova and Lysak 2008). By comparing the macrostructure across the three genomes the closer phylogenetic relationship of the A and C genomes is apparent with a number of linkage groups being homologous along almost their entire lengths (A1/C1, A2/C2, A3/C3) (Fig. 5.2). In contrast there is only one B genome linkage group which appears to be conserved with the A genome but not the C genome (B5/A5), although there is limiting data to suggest that B4 may be aligned completely with A4 and the lower half of C4. These comparisons have suggested direct mechanisms for the reduction of chromosome numbers between the genomes, for example, it appears that A7 and A8 have fused to form B7. When homoeologous regions between the A and the C had been defined previously there was one region of the A genome, the top of A10, which showed no homoeology with the C genome. However, it appears that the organization of A10 is completely conserved on linkage group B8 (Figs. 5.3 and 5.4); it remains ambiguous as to whether the region in question is missing from the C genome.

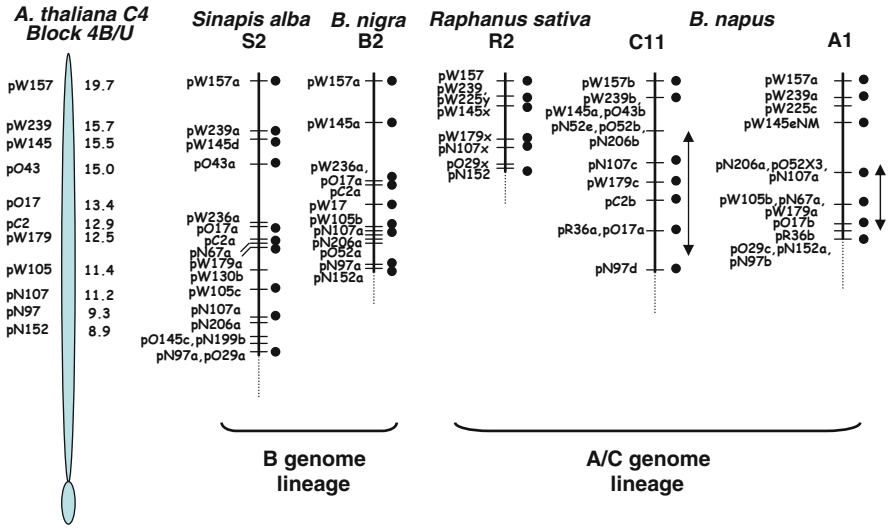

Fig. 5.4 Conservation of ancestral blocks across lineage II species. Presence of block 4B/U in a number of lineage II species which span the divergence of the A and C genomes from the B genome species

5.6.2 Conserved Chromosome Landmarks

Cytological studies in *Brassica* species have been notoriously difficult due to the small size of the chromosomes and it has only been with recent improvements in labeling and imaging technologies that the individual chromosomes can be differentiated in the nucleus (Howell et al. 2008). This has meant that alignment of the linkage maps with the chromosomal karyotype has remained elusive. Comparative mapping has allowed the tentative positions of centromeres to be inferred in the three diploid genomes based on the mapping of markers flanking the *A. thaliana* centromeric regions to conserved collinear regions of the *Brassica* genome (Fig. 5.3). In three instances these regions have been confirmed (C1, C2, and C4) by the genetic mapping of telocentric chromosomes in AAC triploid lines (Kelly and Lydiate, unpublished). These data imply that a significant number of the macro-rearrangements which differentiate the three diploid genomes are the result of chromosome fission and fusion at centromeric and telomeric sites.

5.6.3 Rearrangement Hotspots

Due to the extensive restructuring and the unknown architecture of the progenitor genome it is difficult to accurately determine the number of chromosomal rearrangements which led to the present organization of the *Brassica* diploids. However, careful study of the conserved blocks across the three genomes allows a number of observations. Some linkage groups appeared less susceptible to further rearrangement upon formation and certain block relationships were more stably inherited. The conserved blocks ranged dramatically in size both genetically (*Brassica* species) and physically (*A. thaliana*) with the largest region representing almost half the mapped length of each of three linkage groups (A1, C1, and B2) and equivalent to ~9 Mbp of *A. thaliana* chromosome 4 sequence (almost 50% of the physical length) (Fig. 5.3). In contrast some linkage groups, for example, A9, were a virtual mosaic of blocks, suggesting numerous events in its history. As mentioned above, some rearrangements were observed to be common to all three genomes. The stability of such rearrangements and the relatively conserved chromosome organization of certain linkage groups (e.g., A1/C1, A5/B5/C5) implies an adaptive advantage to the observed structure or a genetic or physical impediment to further restructuring. A number of linkage groups showed clusters of collinear regions that were defined by small genetic regions in *B. napus* but large physical regions in *A. thaliana,* in a number of instances, these regions were co-localized with putative centromeric regions (Fig. 5.3). It has been suggested that regions flanking centromeres are fragile sites in the genome predisposed to rearrangements (Moore et al. 1997, Qi et al. 2006). Sequence-level comparisons across lineage I species provide compelling evidence that the pericentromeric regions although found in apparently conserved regions are prone to dynamic divergence with expansion through insertion of genes, pseudogenes, and repetitive mobile elements (Hall et al. 2006). The fluid nature of these regions and the presence of repetitive elements

could lead to a propensity for chromosomal rearrangements which would be consistent with the alignment of centromeric regions with the endpoints of the ancestral blocks (Schranz et al. 2006).

5.7 Exploiting Comparative Mapping for Trait Analysis

Before the absolute extent of genome conservation between *A. thaliana* and its *Brassica* relatives was determined it had been observed that these oilseeds shared common phenotypes. The simple genome, short generation time, and the availability of significant genetics and genomics resources for *A. thaliana* allowed genes controlling such phenotypes to be identified with relative ease. By identifying allelic variants or manipulating homologues of these genes in *B. napus* it was possible to affect similar traits in the crop as determined in the weed. An early example of this was the association of variant alleles for the *FATTY ACID ELONGASE* (*FAE1*) gene in *B. napus* with reduction in erucic acid in the seed, comparatively the *FAE1 A. thaliana* mutant had reduced levels of long chain fatty acids (Roscoe et al. 2001). The conservation of gene function between species suggested that *A. thaliana* could be a valuable source of candidate genes for traits of agronomic importance in *Brassica* crop species.

With the advent of the comparative mapping data it was possible to associate genomic regions in *Brassica* species underlying advantageous traits with conserved regions in *A. thaliana*. These comparisons allow exploitation of the *A. thaliana* genomic sequence for both the development of targeted markers and the identification of potential candidates controlling the expression of traits of interest (Qiu et al. 2006). Such analyses can be of particular value for the analyses of quantitative trait loci (QTL), where the control of the phenotype is complex being conferred by the presence of a number of loci of varying effect. As well as facilitating the identification of additional markers to saturate the QTL region for fine mapping, the comparative mapping data can indicate where the number of loci controlling a phenotype could be a reflection of the high level of duplication present within the *Brassica* genomes. For example, accumulation of aliphatic glucosinolates in the seed is controlled by at least three QTL loci in *B. napus* (Howell et al. 2003) which are localized to homologous regions of the genome (block C5E/W), suggesting that a duplicate gene family, rather than three unrelated genes, could be manipulated to impact a change.

Based on conserved map positions a number of candidate genes, previously characterized in *A. thaliana*, have now been correlated with QTL loci in *B. napus*, *B.oleracea*, and *B. juncea*, particularly for flowering time, inflorescence morphology, and seed glucosinolate biosynthesis (Osborn et al. 1997, Lan and Paterson 2000, Long et al. 2007, Bisht et al. 2009). The use of comparative mapping was also instrumental in cloning the gene responsible for a dwarf phenotype observed and genetically mapped in *B. rapa* (Muangprom and Osborn 2004). The phenotype was located in the conserved C2A/K block where a DELLA protein involved in gibberellic acid biosynthesis (RGA1) was identified. Furthermore, a mutant allele of the

B. rapa homologue was confirmed to reproduce the dwarf phenotype (Muangprom et al. 2005).

However, the use of comparative mapping data can be misleading in regions where the underlying α duplication, or the most recent duplication event in the history of the *A. thaliana* genome, and the subsequent triplication in the *Brassica* genomes causes difficulties in differentiating between conserved regions (Mayerhofer et al. 2005). In addition, as a consequence of random gene loss in the duplicate regions in both species, the generation of targeted markers can be unpredictable (Town et al. 2006).

5.8 Extending the Comparisons to Related Species

Comparative mapping between *A. thaliana* and its close relatives of lineage I has suggested that the widely studied model genome with its low chromosome number is actually an anomaly among its peers, evolved from a progenitor with a chromosome complement of seven or eight. Alignment of the *A. thaliana* genome with that of *Capsella rubella* (Boivin et al. 2004) and *A. lyrata* (Kuittinen et al. 2004) indicated the reduction in chromosome number was largely the result of chromosome fusions rather than elimination of genomic DNA. In addition, such alignments have shown that while *A. lyrata* and *C. rubella* demonstrate strong collinearity, at least seven major rearrangements, including inversion, chromosomal fusion, and translocation events, are specific to the *A. thaliana* genome (Yogeeswaran et al. 2005). The elucidation of the ancestral collinear blocks in *B. napus* (Parkin et al. 2005) and the painting of those blocks on $n = 7$ and $n = 8$ Brassicaceae species have contributed greatly to our understanding of evolutionary steps in the formation of important *Brassica* species (Schranz et al. 2006, Mandakova and Lysak 2008).

The *Brassica* genus is incredibly diverse in genome content, chromosome number, and morphological form and contains a relatively large number of agronomically important species. Despite this collected wealth of diversity, there is interest in capturing and exploiting traits and allelic variation found among related genera, one such example that has been validated for hybrid development in *B. napus* is the use of the ogura cytoplasmic male sterility system identified in *Raphanus sativa* (Primard-Brisset et al. 2005). To facilitate such applications genetic linkage studies using the core set of RFLP probes sequenced in Parkin et al. (2005) have been initiated in *R. sativa* (Bett and Lydiate 2003), *Sinapis alba* (Nelson and Lydiate 2006), and *Moricandia arvense* (Beschorner and Lydiate, unpublished). Perhaps not surprisingly, preliminary analyses of these data corroborate the existence of an ancestral hexaploid and indicate that the ancestral blocks are conserved across the genera (Fig. 5.4). These and additional studies in species of lineage II, from which the *Brassica* crops evolve, will be necessary to identify conserved block arrangements and elucidate the ancestral karyotype for this lineage (Mandakova and Lysak 2008).

5.9 The Promise of Sequenced Genomes

The genomic sequence of the first *Brassica* species, *B. rapa,* will be available shortly (Yang et al. 2005, Hong et al. 2008). This foundational resource will allow a comprehensive analysis of the relationship between a related crop and weed genome, uncovering at the micro-level the impacts of genome duplication and allowing precise identification of rearrangement endpoints, which could point to the evolutionary mechanisms driving such changes. With the advent of more efficient and cost-effective sequencing technologies it is possible to envision all members of U's triangle being scrutinized at the sequence level. Such analyses will empower the improvement of the constituent crops and cross-species comparisons will provide insights into the evolution of the different crop types.

References

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Axelsson T, Bowman CM, Sharpe AG, Lydiate DJ, Lagercrantz U (2000) Amphidiploid *Brassica juncea* contains conserved progenitor genomes. Genome 43:679–688

Bett KE, Lydiate DJ (2003) Genetic analysis and genome mapping in Raphanus. Genome 46: 423–430

Bisht NC, Gupta V, Ramchiary N, Sodhi YS, Mukhopadhyay A, Arumugam N, Pental D, Pradhan AK (2009) Fine mapping of loci involved with glucosinolate biosynthesis in oilseed mustard (*Brassica juncea*) using genomic information from allied species. Theor Appl Genet 118: 413–421

Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome Res 13:137–144

Bohuon EJR, Keith DJ, Parkin IAP, Sharpe AG, Lydiate DJ (1996) Alignment of the conserved C genomes of *Brassica oleracea* and *Brassica napus*. Theor Appl Genet 93:833–839

Boivin K, Acarkan A, Mbulu RS, Clarenz O, Schmidt R (2004) The Arabidopsis genome sequence as a tool for genome analysis in Brassicaceae. A comparison of the Arabidopsis and *Capsella rubella* genomes. Plant Physiol 135:735–744

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433–438

Cheung WY, Champagne G, Hubert N, Landry BS (1997) Comparison of the genetic maps of *Brassica napus* and *Brassica oleracea*. Theor Appl Genet 94:569–582

Devos KM (2005) Updating the 'crop circle'. Curr Opin Plant Biol 8:155–162

Devos KM, Gale MD (2000) Genome relationships: the grass model in current research. Plant Cell 12:637–646

Ferreira ME, Williams PH, Osborn TC (1994) RFLP mapping of *Brassica napus* using doubled haploid lines. Theor Appl Genet 89:615–621

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95:1971–1974

Hall AE, Kettler GC, Preuss D (2006) Dynamic evolution at pericentromeres. Genome Res 16: 355–364

Henry Y, Bedhomme M, Blanc G (2006) History, protohistory and prehistory of the *Arabidopsis thaliana* chromosome complement. Trends Plant Sci 11:267–273

Hong CP, Kwon SJ, Kim JS, Yang TJ, Park BS, Lim YP (2008) Progress in understanding and sequencing the genome of *Brassica rapa*. Int J Plant Genomics 2008:582837

Howell EC, Kearsey MJ, Jones GH, King GJ, Armstrong SJ (2008) A and C genome distinction and chromosome identification in *Brassica napus* by sequential fluorescence in situ hybridization and genomic in situ hybridization. Genetics 180:1849–1857

Howell PM, Sharpe AG, Lydiate DJ (2003) Homoeologous loci control the accumulation of seed glucosinolates in oilseed rape (*Brassica napus*). Genome 46:454–460

Jenczewski E, Alix K (2004) From Diploids to Allopolyploids: the emergence of efficient pairing control genes in plants. Crit Rev Plant Sci 23:21–45

Kianian SF, Quiros CF (1992) Generation of a *Brassica oleracea* composite RFLP map: linkage arrangements among various populations and evolutionary implications. Theor Appl Genet 84:544–554

Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae). Mol Biol Evol 17:1483–1498

Kuittinen H, de Haan AA, Vogl C, Oikarinen S, Leppala J, Koch M, Mitchell-Olds T, Langley CH, Savolainen O (2004) Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. Genetics 168:1575–1584

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

Lagercrantz U, Lydiate DJ (1995) RFLP mapping in *Brassica nigra* indicates differing recombination rates in male and female meioses. Genome 38:255–264.

Lagercrantz U, Lydiate DJ (1996) Comparative genome mapping in Brassica. Genetics 144: 1903–1910

Lan TH, Paterson AH (2000) Comparative mapping of quantitative trait loci sculpting the curd of *Brassica oleracea*. Genetics 155:1927–1954

Landry BS, Hubert N, Etoh T, Harada JJ, Lincoln SE (1991) A genetic map for *Brassica napus* based on restriction fragment length polymorphisms detected with expressed DNA sequences. Genome 34:543–552

Long Y, Shi J, Qiu D, Li R, Zhang C, Wang J, Hou J, Zhao J, Shi L, Park BS, Choi SR, Lim YP, Meng J (2007) Flowering time quantitative trait loci analysis of oilseed Brassica in multiple environments and genomewide alignment with Arabidopsis. Genetics 177: 2433–2444

Lukens L, Zou F, Lydiate D, Parkin I, Osborn T (2003) Comparison of a *Brassica oleracea* genetic map with the genome of Arabidopsis thaliana. Genetics 164:359–372

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

Mandakova T, Lysak MA (2008) Chromosomal phylogeny and karyotype evolution in x=7 crucifer species (Brassicaceae). Plant Cell 20:2559–2570

Mayerhofer R, Wilde K, Mayerhofer M, Lydiate D, Bansal VK, Good AG, Parkin IA (2005) Complexities of chromosome landing in a highly duplicated genome: toward map-based cloning of a gene controlling blackleg resistance in *Brassica napus*. Genetics 171: 1977–1988

Moore G, Foote T, Helentjaris T, Devos K, Kurata N, Gale M (1995) Was there a single ancestral cereal chromosome? Trends Genet 11:81–82

Moore G, Roberts M, Aragon-Alcaide L, Foote T (1997) Centromeric sites and cereal chromosome evolution. Chromosoma 105:321–323

Muangprom A, Osborn TC (2004) Characterization of a dwarf gene in *Brassica rapa*, including the identification of a candidate gene. Theor Appl Genet 108:1378–1384

Muangprom A, Thomas SG, Sun TP, Osborn TC (2005) A novel dwarfing mutation in a green revolution gene from *Brassica rapa*. Plant Physiol 137:931–938

Nelson MN, Lydiate DJ (2006) New evidence from *Sinapis alba* L. for ancestral triplication in a crucifer genome. Genome 49:230–238

Osborn TC, Kole C, Parkin IA, Sharpe AG, Kuiper M, Lydiate DJ, Trick M (1997) Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. Genetics 146: 1123–1129

Panjabi P, Jagannath A, Bisht NC, Padmaja KL, Sharma S, Gupta V, Pradhan AK, Pental D (2008) Comparative mapping of Brassica juncea and Arabidopsis thaliana using Intron Polymorphism (IP) markers: homoeologous relationships, diversification and evolution of the A, B and C Brassica genomes. BMC Genomics 9:113

Parkin IA, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiate DJ (2005) Segmental structure of the Brassica napus genome based on comparative analysis with *Arabidopsis thaliana*. Genetics 171:765–781

Parkin IA, Sharpe AG, Keith DJ, Lydiate DJ (1995) Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). Genome 38:1122–1131

Parkin IA, Sharpe AG, Lydiate DJ (2003) Patterns of genome duplication within the *Brassica napus* genome. Genome 46:291–303

Paterson AH, Bowers JE, Chapman BA, Peterson DG, Rong J, Wicker TM (2004) Comparative genome analysis of monocots and dicots, toward characterization of angiosperm diversity. Curr Opin Biotechnol 15:120–125

Piquemal J, Cinquin E, Couton F, Rondeau C, Seignoret E, Doucet I, Perret D, Villeger MJ, Vincourt P, Blanchard P (2005) Construction of an oilseed rape (*Brassica napus* L.) genetic map with SSR markers. Theor Appl Genet 111:1514–1523

Primard-Brisset C, Poupard JP, Horvais R, Eber F, Pelletier G, Renard M, Delourme R (2005) A new recombined double low restorer line for the Ogu-INRA cms in rapeseed (*Brassica napus* L.). Theor Appl Genet 111:736–746

Qi L, Friebe B, Gill BS (2006) Complex genome rearrangements reveal evolutionary dynamics of pericentromeric regions in the Triticeae. Genome 49:1628–1639

Qiu D, Morgan C, Shi J, Long Y, Liu J, Li R, Zhuang X, Wang Y, Tan X, Dietrich E et al (2006) A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content. Theor Appl Genet 114:67–80

Roscoe TJ, Lessire R, Puyaubert J, Renard M, Delseny M (2001) Mutations in the fatty acid elongation 1 gene are associated with a loss of beta-ketoacyl-CoA synthase activity in low erucic acid rapeseed. FEBS Lett 492:107–111

Sabhyata B, Madan Singh N, Malathi L (1996) Structural analysis of the rDNA intergenic spacer of *Brassica nigra*: evolutionary divergence of the spacers of the three diploid brassica species. J Mol Evol 43:460–468

Schranz ME, Lysak MA, Mitchell-Olds T (2006) The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends Plant Sci 11:535–542

Slocum MK, Figdore SS, Kennard WC, Suzuki JY, Osborn TC (1990) Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. Theor Appl Genet 80: 57–64

Suwabe K, Morgan C, Bancroft I (2008) Integration of Brassica A genome genetic linkage map between *Brassica napus* and *B. rapa*. Genome 51:169–176

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008) Synteny and collinearity in plant genomes. Science 320:486–488

Tanksley SD, Ganal MW, Prince JP, de-Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160

Town CD, Cheung F, Maiti R, Crabtree J, Haas BJ, Wortman JR, Hine EE, Althoff R, Arbogast TS, Tallon LJ, Vigouroux M, Trick M, Bancroft I (2006) Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell 18:1348–1359

U N (1935) Genomic analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Jpn J Bot 7:389–452

Uzunova M, Ecke W, Weissleder K, Röbbelen G (1995) Mapping the genome of rapeseed (Brassica napus L.). I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content. Theor Appl Genet 90:194–204

Yang TJ, Kim JS, Lim KB, Kwon SJ, Kim JA, Jin M, Park JY, Lim MH, Kim HI, Kim SH, Lim YP, Park BS (2005) The Korea brassica genome project: a glimpse of the Brassica genome based on comparative genome analysis with Arabidopsis. Comp Funct Genomics 6:138–146

Yogeeswaran K, Frary A, York TL, Amenta A, Lesser AH, Nasrallah JB, Tanksley SD, Nasrallah ME (2005) Comparative genome analyses of Arabidopsis spp.: inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana*. Genome Res 15:505–515