



Plant genome evolution: lessons from comparative genomics at the DNA level

Renate Schmidt

Max-Delbrück-Laboratorium in der Max-Planck-Gesellschaft, Carl-von-Linné-Weg 10, 50829 Cologne, Germany; present address: Max-Planck-Institut für molekulare Pflanzenphysiologie, Am Mühlenberg 1, 14476 Golm, Germany (e-mail rschmidt@mpiz-koeln.mpg.de)

Key words: collinearity, comparative mapping, comparative sequence analysis, genome, retroelement, sequence conservation

Abstract

Angiosperm genomes show tremendous variability in genome size and chromosome number. Nevertheless, comparative genetic mapping has revealed genome collinearity of closely related species. Sequence-based comparisons were used to assess the conservation of gene arrangements. Numerous small rearrangements, insertions/deletions, duplications, inversions and translocations have been detected. Importantly, comparative sequence analyses have unambiguously shown micro-collinearity of distantly related plant species. Duplications and subsequent gene loss have been identified as a particular important factor in the evolution of plant genomes.

Angiosperm genomes

Cytogenetic techniques allow insight into genome organisation at the chromosome level. Chromosome numbers for different species have been established by light microscopic analysis of chromosome spreads. In angiosperms, plants with as few as $2n = 4$ (e.g. *Haplopappus gracilis*) and as many as $2n =$ ca. 600 chromosomes (*Voanioala gerardii*) are known (Bennett, 1998). The importance and prevalence of polyploidy in angiosperms has also been recognized by studying karyotypes of different plant species. It has been estimated that 50–70% of flowering plants have experienced chromosome doubling at least once in their evolutionary history (Wendel, 2000). Many of the important crop plants are polyploids (e.g. wheat, rapeseed, potato, cotton).

Different methods can be used to estimate genome size. These include for example DNA reassociation kinetics, nuclear volume measurements and estimations from sampling genomic clone libraries. Microdensitometry of Feulgen-stained nuclei (Bennett and Smith, 1976, 1991; Bennett *et al.*, 1982) and flow cytometry of isolated nuclei stained with propidium iodide

(Arumuganathan and Earle, 1991) were used for extensive surveys. A compilation of 2802 estimates for angiosperm species has shown that haploid genome sizes range over 1000-fold from ca. 0.1 pg to over 125 pg. About 50% of the flowering plants analysed to date have genome sizes between 0.1 and 3.5 pg (Leitch *et al.*, 1998). *Arabidopsis* has one of the smallest genomes observed in higher plants, the analysis of the DNA sequence of the nuclear genome supports a value of 125 Mb (Arabidopsis Genome Initiative, 2000). In contrast, the particularly large genome of *Fritillaria assyriaca* encompasses ca. 120 000 Mb (Bennett and Smith, 1976). Even species belonging to the same family show substantial differences in genome size. In the Poaceae, values of ca. 450, 750, 2500, 5000 and 16 000 Mb have been established for the rice, sorghum, maize, barley and wheat genomes, respectively (Arumuganathan and Earle, 1991).

Reassociation kinetic studies provide an important insight into the complexity of plant genomes. These experiments have unequivocally shown that plant genomes are composed of repeated and low or single-copy DNA sequences. Comparing complexities of large and small plant genomes, it has been estab-

lished that differences in genome size can mainly be attributed to the varying proportion of repeated DNA sequences (Flavell, 1980), although the ploidy level is another factor which affects the size of genomes.

Repeated DNA sequences can be divided into two classes, elements which are organized in tandem arrays and those which show a dispersed distribution in the genome. Transposons are a particular frequent component of the latter category. In the small *Arabidopsis* genome transposable elements account for only 10% of the genome (*Arabidopsis* Genome Initiative, 2000), whereas the 20-fold larger maize genome consists of at least 60–80% of repetitive DNA sequences. The especially abundant retrotransposons are found at a high frequency interspersed with gene sequences in the maize genome and make up more than 50% of the maize genome (SanMiguel *et al.*, 1996; Rabinowicz *et al.*, 1999). In contrast, the roughly 2100 retroelement copies present in the *Arabidopsis* genome, mainly in the centromeric regions, make up less than 10% of the genome (*Arabidopsis* Genome Initiative, 2000).

Hybridization-based comparisons of plant genome organization

Gene coding sequences from closely related plant species show considerable DNA similarity. Hence, it can be tested which proportion of gene coding sequences of a given species cross-hybridizes with DNA sequences in a related species. It cannot be discriminated, however, if a particular sequence is missing from a genome altogether or has diverged to an extent that it does not have sufficient sequence similarity to be detectable in hybridization experiments. Despite this limitation such studies give a good indication of the conservation of gene repertoires in related species.

Markers which reveal restriction site polymorphisms in DNA of different individuals in genomic DNA blot hybridizations (RFLP markers) have been adopted to establish genetic linkage maps for many different plant species. Gene, cDNA or random low-copy DNA sequences are used as RFLP markers. Thus, if a collection of RFLP markers from one species is tested for cross-hybridization to DNA from a related species the similarity of the low-copy sequence repertoires of these two species is assessed.

Species of the Solanaceae (tomato, potato and pepper) share a highly conserved gene repertoire. As many of 46 cDNA clones hybridized to tomato as

well as pepper DNA, regardless of whether the cDNA clones were derived from tomato or pepper (Tanksley *et al.*, 1988). Likewise, nearly all tomato cDNA or genomic clones tested hybridized to potato DNA (Bonierbale *et al.*, 1988). Similar results were obtained if species of the Poaceae were compared in respect to marker repertoire. Of 105 maize RFLP markers tested, only one failed to hybridize to sorghum DNA, however, 15–20 other probes hybridized much more strongly to maize than to sorghum DNA (Hulbert *et al.*, 1990). About 85% of rice, oat and barley cDNA clones analysed showed hybridization to maize DNA (Ahn and Tanksley, 1993). According to these results, RFLP markers derived from one species can be exploited for genetic mapping in related species. If genetic maps are constructed with the same set of RFLP markers for two or more species it is possible to compare the resulting linkage maps. Thus it can be determined whether the order of markers along the linkage groups is conserved in the species studied. This is a very powerful technique to compare the gross chromosomal organization of two or more species, especially since only a limited number of markers is needed for such a comparison.

Genetic mapping, however, limits the resolution of such studies. For example, it will often not be possible to establish the order of physically closely linked markers in an unambiguous way using genetic mapping, especially if small segregating populations are used. Likewise, if sequences are mapped for which several cross-hybridizing sequences exist in the species analysed, it is often not possible to determine whether orthologous loci are compared.

Only a small proportion of genes from a given genome is analysed in comparative genetic mapping studies. Consequently, many tens or hundreds of genes may be present in an interval delimited by a pair of adjacent markers. A more detailed characterization of orthologous regions in related species is necessary to reveal whether local gene order, orientation and spacing are conserved between species. This can be accomplished by comparative physical mapping and sequencing. Such studies require libraries of cloned genomic DNA fragments. For any given cloned genomic DNA fragment the gene content needs to be assessed and compared to the orthologous region in the other species of interest.

In some cases information about the gene content may be available for a particular genomic region in one of the species studied. Then the gene content of the orthologous region in another species can be

assayed using DNA blot experiments of digested high-molecular-weight DNA, which has been separated by pulsed-field gel electrophoresis. Alternatively, extended DNA fibres can be subjected to fluorescent *in situ* hybridization.

Sequence-based comparisons of plant genome organization

The availability of sequence information for a genomic DNA fragment offers unique opportunities for its analysis. The gene repertoire of this particular region can be established. Moreover, the spacing of genes and their orientation relative to each other may be studied. It is also possible to analyse exon/intron structures in detail. Furthermore, it can be examined whether the region shows any hallmarks of repetitive sequences such as mobile elements.

It is immediately obvious that sequence analyses of orthologous regions allow comparisons of unprecedented detail. Sequence alignments indicate which kind of sequences show conservation. Gene repertoire, spacing and order can be unambiguously compared in orthologous regions. Most importantly, comparative sequence analysis identifies the nature of differences in gene arrangements.

The low degree of sequence identity in distantly related species hampers an unambiguous recognition of orthologous sequences using hybridization-based approaches. In contrast, comparisons at sequence level are much more sensitive, even regions from distantly related species can be reliably analysed.

Large-scale duplications in plant genomes

The generation of a genetic linkage map is not only useful for assigning loci to positions on chromosomes. Genetic mapping may also highlight duplicated areas of a genome. If markers are utilized which detect two different loci each in a given genome, it can be studied whether the duplicated sequences are randomly arranged in the genome or whether pairs of loci are found in a collinear pattern. An ordered arrangement of duplicated sequences along pairs of chromosomes points to the common origin of these chromosomal segments. Such a pattern could be the result of a duplication of a chromosome segment or it could indicate the polyploid ancestry of a genome.

The hexaploid bread wheat genome consists of three sets of seven homoeologous chromosomes. RFLP mapping has revealed that the majority of gene sequences are triplicated. A comparison of the chromosome linkage maps has shown that the identity of gene orders on homoeologous chromosomes is only interrupted by few gross chromosomal rearrangements (Chao *et al.*, 1989; Devos and Gale, 1993).

In maize, 28.6% of cloned sequences tested detected more than one fragment on genomic Southern blots. Mapping of these duplicated sequences has revealed that they were arranged in a non-random order. Thirteen pairs of duplicate loci were, for example, present on chromosomes 2 and 7. The order of the loci was roughly the same on both chromosomes and the loci were distributed in chromosomal segments, which spanned more than 50 cM each. Generally, duplicated sequences in maize have been found in an ordered arrangement along pairs of chromosome segments. This pattern of duplicated loci in maize supports the polyploid origin of maize even if the current maize genome does not consist of five pairs of homoeologous chromosomes (Helentjaris *et al.*, 1988).

Similarly, genetic mapping in *Brassica nigra*, *B. oleracea* and *B. rapa* showed that a high proportion of sequences in these genomes is duplicated (Slocum *et al.*, 1990; Song *et al.*, 1991; Truco and Quiros, 1994; Lan *et al.*, 2000). But, only genetic mapping experiments which took advantage of a particularly polymorphic cross in *B. nigra* have disclosed eight chromosomal segments, which are present in three copies each (Lagercrantz and Lydiate, 1996). Collinear regions corresponding to these chromosomal segments were also identified in the *B. oleracea* and *B. rapa* genomes. From these data it was concluded that the three *Brassica* species studied have triplicated genomes and a hexaploid ancestor was proposed (Lagercrantz and Lydiate, 1996).

An example for a duplication, which only involves a particular region of a genome, was found by carrying out genetic mapping experiments in rice. The duplicated segment encompasses the distal ends of the short arms of chromosomes 11 and 12. Clone contig maps of these regions were constructed to allow a detailed study. In the duplicated segments, which spanned ca. 2.5 Mb each, 35 DNA markers were found in a collinear arrangement. Only two of the markers tested appeared to be single-copy sequences, the markers were present on chromosome 11 but not chromosome 12 (Wu *et al.*, 1998).

These examples show that comparative mapping is a powerful tool to discover duplicated segments in a given genome. However, the limited resolution of these experiments does not allow a reliable estimate of the similarity of gene repertoire and arrangement in duplicated segments. Using genetic mapping many markers may for example only be mapped to one of the copies of a particular duplicated segment due to a lack of suitable polymorphisms.

Based on the results of genetic mapping the presence of large-scale duplications has also been suggested for the *Arabidopsis* genome (Kowalski *et al.*, 1994; Grant *et al.*, 2000; Lan *et al.* 2000). Extensive comparative sequence analysis has corroborated these observations. Sequencing of a 400 kb contig on chromosome 4 and comparison of that region to other *Arabidopsis* sequences has shown that for nine out of eleven genes in a 45 kb region, counterparts are present in a conserved arrangement on chromosome 2. The two regions differ by insertion/deletion of several genes and non-coding sequences are not conserved. Therefore it has been concluded that this duplication was ancient (Terry *et al.*, 1999). Upon availability of the complete sequences for chromosomes 2 and 4 it became apparent that the region described is only a small part of a much larger duplication. The copy on chromosome 2 spans 4.6 Mb, and 430 out of 1100 genes are in common to the regions on chromosomes 2 and 4. Apart from megabase-scale rearrangements gene order is preserved in the two segments (Lin *et al.*, 1999; Mayer *et al.*, 1999; Bancroft, 2000). The characteristics that have been established for this duplication can also be traced in other duplicated segments of the *Arabidopsis* genome (Lin *et al.*, 1999; Blanc *et al.* 2000; Rossberg *et al.*, 2001; Figure 1A). Based on the analysis of the genome sequence it has been estimated that evidence for ancient duplications is found for ca. 60% of the genome (Blanc *et al.*, 2000; Paterson *et al.*, 2000; Arabidopsis Genome Initiative, 2000; Vision *et al.*, 2000).

Comparative genomics between closely related species

Numerous comparisons of genome structure using genetic mapping between closely related species have been carried out. A special emphasis has been on species of the Poaceae and the Brassicaceae, although several studies have been performed for members of the Solanaceae and legumes (Schmidt, 2000).

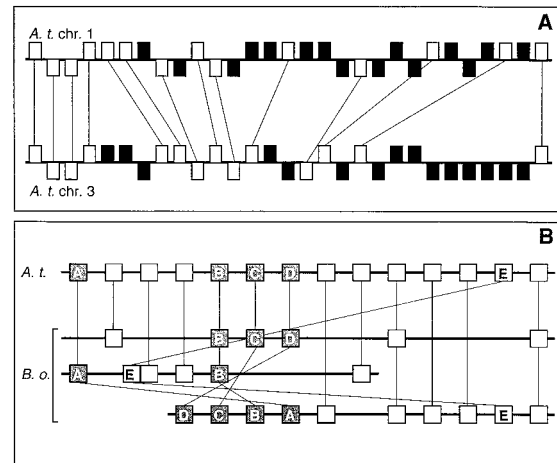


Figure 1. Duplicated chromosome segments differ in gene content. A. The gene arrangement of a region mapping to *A. thaliana* chromosome 1 is compared to its counterpart on chromosome 3. Genes in common to both regions are shown as white boxes. Lines connect homologous sequences. Black boxes represent genes unique to the chromosome 1 or chromosome 3 region, respectively. The location of a box relative to the sequence drawn as a line indicates the direction of transcription (Rossberg *et al.*, 2001). B. Comparison of a region of the *A. thaliana* genome and three corresponding homoeologous segments from *B. oleracea*. A square indicates the presence of a gene in a particular genomic region. Lines connect homologous sequences. Gene arrangements indicative of a translocation (E) and an inversion (A–D) are shaded light and dark grey, respectively (O'Neill and Bancroft, 2000).

Comparative mapping has generally revealed collinear chromosomal segments in closely related plants, albeit of varying size. In some cases entire chromosomes show collinearity. For example, only five chromosomal inversions have to be inferred to explain differences in marker organization between the twelve tomato and potato chromosomes (Tanksley *et al.*, 1992). In contrast, other comparative studies revealed collinear regions spanning only few centimorgans. Such a pattern was, for example, described for the *A. thaliana* and *Brassica nigra* genomes. One has to assume about 90 chromosomal rearrangements since the divergence of these species to explain the observed pattern of collinear segments which span on average 8 cM (Lagercrantz, 1998). The rate at which chromosomal rearrangements have taken place in the *A. thaliana* and *B. nigra* genomes is far higher than values that have been reported for the Poaceae (Paterson *et al.*, 1996). Comparative mapping between the *Arabidopsis* and *Capsella* genomes revealed much larger collinear segments than those observed for *A. thaliana* and *B. nigra* (Acarkan *et al.*, 2000). Although *Arabidopsis* and *Capsella* diverged more re-

cently than *A. thaliana* and *B. nigra*, this difference does not fully account for the numerous rearrangements reflected in the comparative maps of *Arabidopsis* and *Brassica*. Rather these results indicate that different rates of chromosomal rearrangements are observed in the Brassicaceae if different species pairs are studied.

Comparative genetic mapping between species of the Poaceae revealed extensive genome collinearity even if species were compared which diverged as long as 60 million years ago. Moreover, genome sizes of some of the species studied varied as much as 40-fold (Gale and Devos, 1998). It was possible to establish that a limited number of rice linkage segments is sufficient to describe the marker arrangement on the twelve rice, seven wheat and ten maize chromosomes. In accordance with the polyploid origin of maize each of the rice linkage segments was found to correspond to two different maize chromosomes (Moore *et al.*, 1995). Based on the concept of conserved linkage segments multiple alignments of chromosome maps are possible and a comparative map including the genomes of foxtail millet, oats, pearl millet, maize, rice, sugarcane, sorghum and Triticeae was developed (Gale and Devos, 1998). Interestingly, despite the close taxonomic relationship of pearl and foxtail millet, comparative mapping revealed a large number of gross chromosomal rearrangements. A comparison of the millet and rice genomes indicated that most of these structural changes very likely took place in pearl millet (Devos *et al.*, 2000).

Borders that delimit conserved linkage blocks can often be aligned with the sites of centromeres, telomeres and nucleolar organizer regions. This has been shown in comparisons involving different species of the Poaceae, Brassicaceae and Solanaceae, respectively (Moore *et al.*, 1997; Lagercrantz *et al.*, 1998; Livingstone *et al.*, 1999). Aligned chromosome maps allow the identification of markers derived from various species for a given genomic region. This approach is especially useful if large numbers of markers are needed for fine-scale mapping and map-based cloning experiments. Minor deviations from overall collinearity are, however, frequently detected if more refined mapping experiments are carried out. For example, rice chromosome 9 is collinear with the consensus map for group 5 chromosomes of wheat, but a non-syntenic region was pinpointed. Probes from this region of wheat chromosome 5 map to rice chromosomes 2, 8 and 11 (Foote *et al.*, 1997). Likewise, a detailed comparative study of the *Rpg1* genomic regions

in rice and barley provided evidence for a translocation which disrupts collinearity (Kilian *et al.*, 1995; Kilian *et al.*, 1997). Comparing the marker arrangement on corresponding linkage groups of *A. thaliana* and *Capsella* revealed that two of the *Arabidopsis* markers located on chromosome 4 were not present in the *Capsella* genome. For another marker, which represents a single-copy gene on *A. thaliana* chromosome 4 two unlinked loci were found in *Capsella* (Acarkan *et al.*, 2000). These examples show that the genome arrangement in related species is not only distinguished by large-scale chromosomal rearrangements but that many small structural changes, such as deletions/insertions, duplications and translocations of gene sequences are frequently observed.

From results of comparative genetic and physical mapping experiments it is difficult to assess to what extent genome structure is conserved in two species. Lack of polymorphisms often does not allow one to analyse the map positions of all loci corresponding to a particular marker in the species of interest. Thus it may be impossible to determine whether a marker in a non-collinear position provides evidence for a rearrangement or whether a paralogous sequence has been mapped. Likewise, on the basis of hybridization experiments it is difficult to establish whether a sequence is absent from a given genome or whether lack of hybridization is due to a high degree of sequence divergence. Therefore, it is important to investigate collinearity of genomes at the sequence level.

So far, only few comparative sequence analyses have been carried out but they indicate that microstructure might not be as conserved as the gross chromosomal organization (Bennetzen, 2000). Rice and sorghum diverged about 50 million years ago, nevertheless complete micro-collinearity has been established for the *sh2/al* region. In addition to the *sh2* and *al* genes, a putative transcriptional regulatory gene has been identified in the corresponding regions of both genomes. Evolutionarily conserved sequences in the chromosome segments analysed correspond to genes, with intron sequences evolving at a much faster rate than exons (Chen *et al.*, 1998). In both rice and sorghum the *sh2* and *al* genes are separated by ca. 19 kb, whereas in maize the two genes are 140 kb apart (Chen *et al.*, 1997). Comparing the *adh* regions of maize and sorghum, it has also been found that the distances between genes are different in both species. Nine genes were pinpointed in the 225 kb sequence of the maize *adh* region and these are present in the same order in sorghum. Five additional genes were de-

tected in the *adh* region of sorghum, although at 80 kb it is much smaller than the corresponding maize segment. Three of these five sorghum genes are flanked by genes, which are present in sorghum and maize. Hybridization studies have shown that these three genes are located elsewhere in the maize genome. The increased size of the maize *adh* region compared to the one of sorghum is due to the presence of many retrotransposons (Tikhonov *et al.*, 1999).

Earlier studies had already indicated that many of the repetitive elements do not cross-hybridize between maize and sorghum (Hulbert *et al.*, 1990). This feature can be exploited to identify gene sequences in complex genomes, since only such sequences will cross-hybridize with DNA from related species (Avramova *et al.*, 1996). A comparison of the 22 kDa α -zein cluster in maize and the corresponding regions in sorghum and rice has also revealed that the presence of repetitive elements located amidst gene sequences explains the observed size differences in intergenic regions of these species (Messing and Llaca, 1998).

Comparative sequence analysis of barley and rice regions containing four conserved genes has revealed more retroelement sequences in the barley than in the rice segment. The four genes are present in the same orientation in the rice genome, but in the barley region one of the genes is inverted in respect to its neighbours. Another difference distinguishing the gene arrangement in both species is that one gene is present in two tandemly arranged copies in the barley genome, whereas the rice region harbours a single copy (Dubcovsky *et al.*, 2001; Figure 2C).

The results of the described micro-collinearity studies comparing regions of the barley, maize, sorghum and rice genomes suggest that the sizes of intergenic regions are correlated with genome size (Messing and Llaca, 1998; Tikhonov *et al.*, 1999; Dubcovsky *et al.*, 2001). In contrast, receptor-like kinase genes were found tightly clustered, not only in the small rice genome but also in the much larger wheat and barley genomes. A detailed comparison of the regions harbouring receptor-like kinase genes in the wheat, barley, rice and maize genomes has also revealed the important role of duplications and other small-scale rearrangements in plant genome evolution (Feuillet and Keller, 1999).

Sequence analysis coupled with genetic and/or physical mapping experiments has also been exploited for detailed comparative studies. For the 340 kb of DNA sequence around the *Adh1* and *Adh2* loci of rice, the presence of 33 genes was predicted. Only

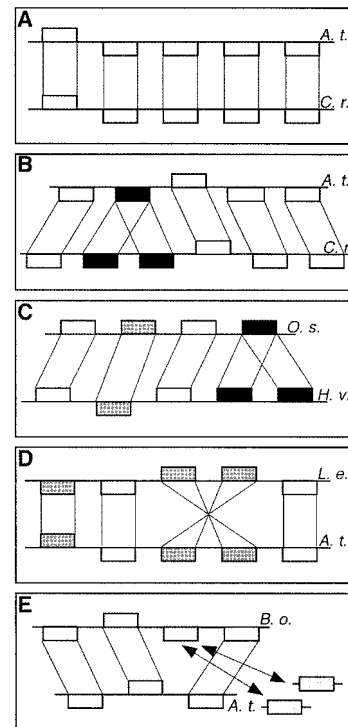


Figure 2. Comparison of gene arrangements in orthologous regions. Gene sequences are shown as boxes and lines connect homologous genes. Those genes, which indicate deviations from micro-collinearity are shaded grey or black. The different directions of transcription are shown by the locations of the boxes relative to the sequence drawn as a line. A. A region of *A. thaliana* (*A. t.*) chromosome 1 is completely collinear with its counterpart in *C. rubella* (*C. r.*) (Rossberg *et al.*, 2001). B. Orthologous regions of the *A. thaliana* (*A. t.*) and *C. rubella* (*C. r.*) genomes show evidence for a gene duplication (Acaran *et al.*, 2000). C. Corresponding segments of the barley (*H. v.*) and rice (*O. s.*) genomes differ by an inversion and a gene duplication (Dubcovsky *et al.*, 2001). D. A comparison of orthologous regions of the *Arabidopsis* (*A. t.*) and tomato (*L. e.*) genomes reveals two inversions involving one and two genes, respectively (Rossberg *et al.*, 2001). E. A translocation differentiates corresponding segments of the *Brassica oleracea* (*B. o.*) and *Arabidopsis* (*A. t.*) genomes. Grey shading and arrows highlight the genes involved in the translocation (Quiros *et al.*, 2001).

five out of thirteen rice genes tested cross-hybridized with maize DNA. Sequence information for the *adh2* region in maize would be a necessary prerequisite to determine whether the lack of cross-hybridization to maize DNA for 8 of the 13 rice genes indicates a difference in gene repertoire of these species or a low degree of sequence conservation. Genetic mapping studies have shown that four of the cross-hybridizing genes are located in a chromosome segment on maize chromosome 4, but the *adh1* gene maps to maize chromosome 1 (Tarchini *et al.*, 2000). This is indicative of a translocation of the *adh1* gene. The complete lack

of collinearity between the *adh1* region of maize and the segment of the rice genome which carries the orthologue of *adh1* is consistent with this assumption (Tikhonov *et al.*, 1999).

In the Brassicaceae the annotated sequence of the *Arabidopsis* genome has been exploited for comparative mapping studies in a very similar fashion as described for the comparative study of Tarchini *et al.* (2000). Nineteen different gene sequences located in a 222 kb segment of *A. thaliana* chromosome 4 were chosen for a comparative analysis with *Brassica oleracea*. For 9 of the 19 genes duplicated copies are present in collinear arrangement on *Arabidopsis* chromosome 5. The 19 different gene sequences were used as probes to identify *B. oleracea* BAC clones harbouring homologous sequences. Seven different BAC contigs were established. Three contigs corresponded to the *Arabidopsis* chromosome 5 region. All three *B. oleracea* regions showed collinearity with the *Arabidopsis* counterpart, but in any one of the triplicated *B. oleracea* segments one or several of the genes located in the *Arabidopsis* region were missing. Only the gene content of all three *B. oleracea* contigs taken together equalled that of *Arabidopsis*. A comparison of the *B. oleracea* contigs with their counterpart on *Arabidopsis* chromosome 4 revealed very similar results; in addition, evidence for a translocation and an inversion was detected (O'Neill and Bancroft, 2000; Figure 1B). The results of this microsynteny study are consistent with the proposed triplicated nature of the *B. oleracea*, *B. nigra* and *B. rapa* genomes (Lagercrantz and Lydiate, 1996). Consequently, in the amphidiploid oilseed rape genome up to six different copies correspond to a particular *A. thaliana* segment (Parkin *et al.*, 1995; Bohuon *et al.*, 1996; Scheffler *et al.*, 1997; Cavell *et al.*, 1998). The *A. thaliana* region carrying the *GTP*, *RPM1* and *M4* genes is, for example, represented six times in the *B. napus* genome. Two of the *B. napus* loci contain all three genes, whereas in the remaining four loci the *RPM1* gene appears to be deleted (Grant *et al.*, 1998). Other comparative studies between the *Brassica* and *Arabidopsis* genomes have also revealed evidence for differences in gene content in homoeologous *Brassica* segments (Sadowski *et al.*, 1996; Sadowski and Quiros, 1998). Thus, deletions appear to occur very frequently in multiplied regions of a genome.

A comparison of a segment of the *B. campestris* genome harbouring the self-incompatibility genes and the corresponding region of the *A. thaliana* genome has revealed extensive collinearity at the sub-

megabase scale. Nevertheless, evidence for a small inversion, translocations and gene deletions/insertions was detected. Three out of 21 *A. thaliana* genes mapping to a 275 kb region did not cross-hybridize with *B. campestris* DNA and the *B. campestris* *SLG* and *SRK* genes were not found in the *A. thaliana* region (Conner *et al.*, 1998).

For the region corresponding to the *Rps2* region of *A. thaliana* chromosome 4, a collinear segment on chromosome 4 of *B. oleracea* was identified. However, in the *Brassica* region an additional gene is found, which is homologous to genes located on *Arabidopsis* chromosomes 2 and 5. Thus, this microsynteny study highlighted a translocation (Quiros *et al.*, 2001; Figure 2E).

Complete micro-collinearity was observed in comparative studies of the closely related species *Arabidopsis thaliana* and *Capsella rubella*. For two 30 kb regions located on *A. thaliana* chromosomes 1 and 4, respectively, it has been shown that gene order and orientation is identical in both species (Acarkan *et al.*, 2000; Rossberg *et al.*, 2001; Figures 2A and 2B). A single difference was detected; one out of the eleven genes studied was tandemly duplicated in *C. rubella* but not in *A. thaliana* (Figure 2B). As judged by the pattern of amino acid exchanges the duplication of the gene took place in *Capsella* after the divergence of *Arabidopsis* and *Capsella* (Acarkan *et al.*, 2000). Corresponding chromosome segments in *Arabidopsis* and *Capsella* are very similar in size (Acarkan *et al.*, 2000; Rossberg *et al.*, 2001). In comparisons between genomic regions of *Arabidopsis* and *Brassica* it depended on the particular segments analysed whether similar-sized regions were observed in both species or whether an increase in size was noted for the *Brassica* segment when compared to the *Arabidopsis* counterpart (Sadowski *et al.*, 1996; Conner *et al.*, 1998; Grant *et al.*, 1998; Sadowski and Quiros, 1998; Jackson *et al.*, 2000; O'Neill and Bancroft, 2000).

Comparative genomics between distantly related species

The low degree of sequence identity, as is generally found for orthologous genes in distantly related species, does in many cases not allow the unambiguous recognition of orthologues by hybridization-based techniques. Only highly conserved gene sequences are suitable markers for cross-hybridization experiments between species belonging to different plant families.

However, many conserved gene sequences are of limited use for comparisons since they belong to gene families. This severely restricts the number of markers available for such experiments and, as a result, conserved linkages might escape detection.

The number of chromosome rearrangements observed in comparative mapping studies between species pairs belonging to different families was compiled. An average rate of $0.14 (\pm 0.06)$ structural mutations per chromosome per million years of divergence was calculated and it was estimated that 43–58% of chromosomal tracts of ≤ 3 cM should remain collinear over a period of 130–200 million years. According to these predictions even monocotyledonous and dicotyledonous species which diverged about 130–200 million years ago should share small collinear chromosome segments (Paterson *et al.*, 1996). Pairs of genes linked at ≤ 3 cM in crucifer plants were used for genetic mapping in cotton and sorghum. In some of the cases studied, linkage of the gene pairs has also been established for sorghum and cotton; however, the distances between the linked markers in sorghum and cotton were often much larger than 3 cM (Paterson *et al.*, 1996). It would be very interesting to assess by statistical analysis to what extent the observed pattern of linkages are expected to be seen by random chance in distantly related species.

Conserved linkage arrangements in distantly related dicotyledonous plants were reported in a study by Grant *et al.* (2000). Soybean linkage group A2, for example, showed significant synteny over its entire length with *Arabidopsis* chromosome 1 and only a limited number of chromosomal rearrangements had to be assumed to explain differences in map order.

Comparative studies between the *Arabidopsis* and rice genomes were undertaken to assess the degree of collinearity between monocotyledonous and dicotyledonous species. Devos *et al.* (1999) analysed regions of *Arabidopsis* chromosome 1. BLAST searches identified rice ESTs with homology to *Arabidopsis* gene sequences located on several BAC clones. A subset of 33 EST sequences, putatively orthologous to the genes derived from *Arabidopsis* chromosome 1, was then used for genetic mapping in rice. Loci corresponding to these EST sequences were found on 10 of the 12 rice chromosomes. For some pairs of *Arabidopsis* locus linkage has also been detected in rice, but generally with the approach chosen conservation of gene order was not detectable between *Arabidopsis* and rice.

Van Dodeweerd *et al.* (1999), with a similar strategy, identified a conserved segment spanning 200–

300 kb in the rice and *Arabidopsis* genomes. Gene predictions derived from a 252 kb region of *A. thaliana* chromosome 4 were used for BLAST searches to reveal homologous rice EST sequences. In total, 24 different ESTs were obtained which putatively represented orthologues of the *Arabidopsis* genes. Among these ESTs were two sequences, that had been used as RFLP markers in rice, the two markers mapped adjacent to each other on rice chromosome 2. A clone contig was established which spanned the region of the rice genome harbouring these two RFLP markers. Three other ESTs, which had been identified in the BLAST analysis, have also been mapped to this clone contig. The order of the five different genes in *Arabidopsis* was distinguished from that in rice by a single inversion. Moreover, the *Arabidopsis* and rice segments showed a different gene content, the conserved framework of genes was interspersed with non-conserved genes. Subsequent analysis showed that the remaining 19 EST sequences were mapping elsewhere in the rice genome.

Further support for a conserved linkage of genes in distantly related species comes from a comparison of a 33 kb rice contig sequence containing five different genes and the sequence of the *Arabidopsis* genome. Two rice genes separated by 16 kb show amino acid similarity with two genes in a similarly sized region mapping to the long arm of *Arabidopsis* chromosome 4. Whereas one additional putative gene is found between the two rice genes, three predicted genes are located between the *Arabidopsis* homologues. Two other putative genes located in the 33 kb rice region also show similarity to genes located in a different region of *A. thaliana* chromosome 4. In this case the orientation of the genes with respect to each other is different in the two species (Han *et al.*, 1999). In contrast, homologues of four genes present in conserved regions of the rice and barley genomes were found to be dispersed in the *Arabidopsis* genome (Dubcovsky *et al.*, 2001).

The comparative analysis of genomic regions in rice and *Arabidopsis* revealed the difficulties and pitfalls of collinearity studies in distantly related species. The unambiguous identification of orthologues in incomplete sequence databases has been identified as the major limitation encountered in such experiments (Devos *et al.*, 1999). However, the completion of the *Arabidopsis* genome sequencing project (*Arabidopsis* Genome Initiative, 2000) and that of rice well advanced (Barry, 2001; Yuan *et al.*, 2001) offers the unique opportunity to assess genome collinearity in

a comprehensive way at the DNA sequence level. This will clarify whether a framework of conserved genes (van Dodeweerd *et al.*, 1999) can be generally observed. The results of the first comparative sequence analyses between distantly related dicotyledonous species show the strength of such an approach.

A pattern of complex relationships was revealed in a study of a 105 kb segment of tomato chromosome 2 and related regions in *Arabidopsis*. The portion of the tomato genome showed conservation of gene content and order with four different segments in the *A. thaliana* genome. The gene repertoire and order observed in the related *Arabidopsis* regions is compatible with assuming at least two consecutive rounds of duplications of an ancestral segment in the *Arabidopsis* lineage followed by extensive loss of genes in duplicated regions (Ku *et al.*, 2000).

Further support for micro-collinearity between the *Arabidopsis* and tomato genomes has also been found in another study. Five different genes were identified in the 57 kb *Lateral suppressor* region of tomato chromosome 7. All five genes have homologues in a region mapping to *Arabidopsis* chromosome 1, which encompasses ca. 30 kb. The arrangement of the five genes in tomato is distinguished from that in *Arabidopsis* by two inversions (Rossberg *et al.*, 2001; Figure 2D).

Tomato and *Arabidopsis* are representative of two major clades of the eudicots (Soltis *et al.*, 1999). In accordance with the results of micro-collinearity studies carried out for tomato and *Arabidopsis* extensive conservation of genome microstructure might also be detectable if genomic regions derived from other dicotyledonous plants are compared. This hypothesis could be tested if sequence information for many different genomic regions would be generated for various dicotyledonous (and monocotyledonous) plants and compared to the sequence of the *Arabidopsis* genome.

Using comparative genome analysis for gene structure predictions

A comparative sequence analysis of the *sh2* and *al* genes from rice, maize and sorghum revealed that exon sequences are considerably more conserved than intron sequences. Interestingly, different rates of divergence for introns are observed in the *sh2* and *al* genes of maize and sorghum, despite a tight linkage of these loci in both genomes. For maize and sorghum coding sequences a high degree of sequence

identity was found, whereas the rice genes are considerably more diverged than their counterparts in maize and sorghum. The exon sequences of the *sh2* gene of maize are, for example, 95% identical to the sorghum homologue while, in contrast, the *Sh2* gene in rice shares 82% and 83% identity with the corresponding genes in maize and sorghum, respectively. This finding is consistent with the divergence times reported for these species. The rice lineage separated from the one of maize and sorghum about 50 million years ago, whereas maize and sorghum diverged 15–20 million years ago (Chen *et al.*, 1998). Aligning *A. thaliana* cDNA sequences with genomic DNA sequences of *A. thaliana* and *C. rubella* revealed not only a high degree of sequence identity for exon sequences at the nucleotide level but also suggested the conservation of number and position of intron sequences in both species. In contrast, the sizes of introns and their sequences vary, although stretches of sequence identity can also be found in intron sequences. Likewise, in intergenic regions no overall sequence homology is found between *Arabidopsis* and *Capsella* genomic DNA sequences (Acarkan *et al.*, 2000).

Evaluation of gene prediction software revealed that gene modelling merely based on gene prediction programs needs further improvement (Pavy *et al.*, 1999). For example, in four out of nine cases analysed the exon/intron structure of a predicted gene differed from that deduced from alignments of *Arabidopsis* cDNA and genomic DNA sequences (Acarkan *et al.*, 2000; *Arabidopsis* Genome Initiative, 2000; Rossberg *et al.*, 2001). Interestingly, aligning such most probably incorrectly predicted *Arabidopsis* coding sequences with *Capsella* genomic DNA sequences did not yield a conserved gene structure. This is in contrast to the results obtained if *Arabidopsis* cDNA sequences are compared to *Arabidopsis* and *Capsella* genomic DNA sequences. Consequently, alignments of genomic DNA sequences of orthologous regions from related species may be exploited to improve gene structure predictions by taking into account conservation of exon length and sequences.

In the segmental duplications of the *Arabidopsis* genome, non-coding sequences are not conserved (Terry *et al.*, 1999; Blanc *et al.*, 2000). Blanc *et al.* (2000) proposed to use conserved exon sequences of duplicated genes as a tool for improvement of gene structure predictions.

Conservation of exon/intron structures has even been seen in sequence alignments of orthologous gene sequences from tomato, *Arabidopsis* and *Capsella*. In-

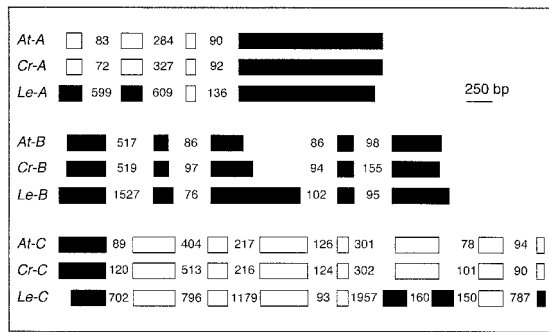


Figure 3. Comparison of exon/intron structures of orthologous genes. The three different coding sequences shown (A, B, C) are present in collinear segments of the *Arabidopsis* (At), *Capsella* (Cr) and tomato genomes (Le) (Rossberg *et al.*, 2001). The regions between start and stop codons are indicated. Exon sequences are displayed as boxes and to scale, for intron sequences the sizes are given in bp. Exons which are identical in length in all three species are represented as white boxes. Grey shading indicates that the exons are identical in size in two of the analysed species.

trons of two different tomato genes were on average twofold larger than their counterparts in *Arabidopsis* or *Capsella*, whereas differences in exon length are essentially restricted to the 5' and 3' regions of the genes. Number of exons was conserved, although for one gene an additional intron has been identified in the tomato copy. One of the five genes analysed in tomato, *Arabidopsis* and *Capsella*, however, showed remarkably different exon sizes (Figure 3). Interestingly, this coincides with a less pronounced level of sequence identity. Most of the sequence comparisons carried out for different gene sequences in *Arabidopsis* and *Capsella* revealed identities of >90% for exon sequences at the nucleotide level, for this particular gene which belongs to the WRKY family of transcription factors a value of ca. 80% was determined. The tomato gene harbours two WRKY domains, whereas in the *Arabidopsis* and *Capsella* genes only the C-terminal domain is present. This difference in domain structure partly accounts for the differences in exon length (Rossberg *et al.*, 2001).

For a gene encoding a putative transcription factor a very similar observation was made; the gene encodes a protein of 895 amino acids in sorghum, whereas the rice protein consists of 1070 amino acids. A putative zinc finger motif present in the rice gene is absent from the sorghum sequence. The differences in predicted protein sizes for the *A1* homologues in sorghum, rice and maize can mainly be attributed to variations in stop codon location at the C-termini of the putative peptides. The *A1* genes in sorghum and maize har-

bour in comparison to the orthologous gene in rice an additional intron (Chen *et al.*, 1998).

Gene structure was found to be largely conserved between rice and barley orthologues. Most of the differences in exon length were confined to the 5' and 3' end of the four genes analysed. Interestingly, a non-canonical splice site in one of the introns was also conserved. These gene structures have been successfully aligned with those of the *Arabidopsis* homologues. Many of the exons analysed proved to be identical in length in all three species. Most notably, these alignments also suggested modifications to *Arabidopsis* gene predictions. *Arabidopsis* exon/intron structures have been deduced for three out of four genes analysed which show a higher overall similarity to the rice and barley genes than the unmodified gene predictions (Dubcovsky *et al.*, 2001). These results clearly demonstrate the utility of comparative sequence analysis for improvement of gene structure predictions.

Detailed comparisons of orthologues from different species may also provide evidence for pseudogenes (Feuillet and Keller, 1999).

Comparative genomics as a tool for gene isolation

Comparative mapping experiments have revealed extensive genome collinearity at the gross chromosomal level between plant species belonging to the same family. In contrast, the results of micro-collinearity suggest a high frequency of small-scale rearrangements, such as deletions/insertions, duplications, inversions and translocations. Although micro-collinearity may be disturbed by duplications of gene sequences and inversions covering one or several genes, this kind of structural alterations does not impose major limitations on using comparative mapping strategies for gene cloning and fine-scale mapping of monogenic or polygenic traits.

The large genome sizes of many important crop plants render map-based cloning experiments especially difficult. Thus, it is attractive to advance such experimental strategies in species with large genomes by exploiting comparative mapping with a related species, which is characterized by a small genome. Due to the numerous deviations seen in microsynteny studies it is nevertheless advisable that the locus to be cloned from a species of interest is covered by a clone contig derived from this plant. Comparative maps are very useful resources to identify many different markers from a variety of species for a given genomic

region. Especially fine-scale mapping or map-based cloning experiments in plants with large genomes may benefit from this. Synteny in the vicinity of *rpg4* was investigated using rice and barley molecular markers as well as clone libraries established from genomic DNA of these species. This approach was successful in delimiting the position of the *rpg4* locus physically and genetically (Druka *et al.*, 2000).

Triticum aestivum has a very large genome of ca. 16000 Mb; moreover, the presence of three highly similar genomes render map-based cloning experiments more difficult. Thus, the extensive collinearity between chromosome 1A^m of the diploid wheat *Triticum monococcum* and chromosome 1A of *Triticum aestivum* (Dubcovsky *et al.*, 1995) was exploited to reduce the complexity of analysis. A physical contig encompassing 450 kb has been established in *T. monococcum*; this region is collinear with the segment of the bread wheat genome which spans the leaf rust resistance locus (Stein *et al.*, 2000).

Correspondence of quantitative traits across different species has been inferred from results of comparative mapping (Lin *et al.*, 1995; Paterson *et al.*, 1995). For example, loci controlling shattering of the inflorescence could be mapped to orthologous regions of foxtail millet, maize, sorghum and rice chromosomes (Paterson *et al.*, 1995; Devos and Gale, 2000).

Homologues of the *Arabidopsis thaliana* *GAI* gene encoding a gibberellin response modulator were cloned from maize, rice and wheat. Using comparative mapping, a gene controlling a key trait in several species has been identified. The *Rht1* gene of wheat and the *D8* gene of maize map to homoeologous chromosome segments. Gene isolation and characterization confirmed their orthology. Mutations in the N-terminal region of the encoded proteins cause reduced response to gibberellin and dwarf phenotypes (Peng *et al.*, 1999).

It can be very difficult to define orthology of genes derived from distantly related species unambiguously, especially if gene families with many members are investigated. But by combining sequence information with micro-collinearity data, orthologous sequences for a member of the rapidly evolving WRKY family of transcription factors have been identified in the *A. thaliana*, *C. rubella* and tomato genomes (Rossberg *et al.*, 2001).

The well-characterized *Arabidopsis* genome together with the extensive collinearity seen in species of the Brassicaceae present unique opportunities for the identification of candidate genes encoding economi-

cally relevant traits in *Brassica*. Reciprocal mapping experiments are carried out to correlate *Brassica* loci of interest with *Arabidopsis* candidate genes.

Putative candidate genes from *Arabidopsis* can serve as molecular markers on suitable segregating populations of *Brassica*. This will show whether any of the loci detected by these probes show cosegregation with the *Brassica* locus of interest. Detailed information on many *Brassica* genomes is available, thus in many cases molecular markers will be available in the vicinity of a particular trait to be studied. These molecular markers can be used for genetic mapping in *Arabidopsis* to identify the corresponding region. For any *Brassica* marker, which represents exon sequences there is a high likelihood that an alignment with the sequence of the *A. thaliana* genome will immediately reveal corresponding genes and their map positions. The mapping of several closely linked *Brassica* marker sequences onto the sequence maps of the *Arabidopsis* chromosomes should pinpoint in most cases a corresponding segment in *A. thaliana*. The annotated *Arabidopsis* sequence can then be used as a tool to refine the positioning and ultimately identify the locus of interest in *Brassica*. The control of flowering time in *Brassica*, for example, is studied by using information about *Arabidopsis* genes that have been implicated in this mechanism (Lagercrantz *et al.*, 1996; Osborn *et al.*, 1997; Bohuon *et al.*, 1998; Lan and Paterson, 2000; Kole *et al.*, 2001). Comparative mapping has identified an oilseed rape homologue of the *Arabidopsis* *CURLY LEAF (CLF)* gene as a candidate for the petal-less flower trait in *B. napus* (Fray *et al.*, 1997). Similarly, homologues of the *Arabidopsis* fatty acid elongase (*FAE1*) gene have been correlated with two loci controlling erucic acid content in oilseed rape (Fourmann *et al.*, 1998).

Pathogen resistance gene homologues were frequently found in non-syntenic map positions in different grasses (Leister *et al.*, 1998). Recently, *Arabidopsis* ESTs and *Brassica* sequences with homology to cloned plant resistance genes were mapped in *B. napus* to provide a source of candidate-resistance genes for *B. napus*. An integration of this information with the map positions of disease resistance loci that have been placed on the oilseed rape genome can now be pursued (Sillito *et al.*, 2000). This will clarify whether a rapid reorganization of disease resistance loci is also observed in the Brassicaceae. For the *Rpm1* and *Rps2* genes it has already been shown that they reside in collinear positions in the *Arabidopsis* and *Brassica* genomes (Grant *et al.*, 1998; Quiros *et al.*, 2001).

The sequence analysis of the *Arabidopsis* (Arabidopsis Genome Initiative, 2000) and rice genomes (Barry, 2001; Yuan *et al.*, 2001) provides a vast resource of gene sequences suitable for genetic and physical mapping experiments. This can be exploited to study a particular genomic region in related species in detail, but markers derived from *Arabidopsis* or rice may also provide an important contribution to establish genome-wide clone contig maps for closely related species. For example, the information from the rice physical map has been recognised as a powerful resource for advancing a contig map of the sorghum genome (Draye *et al.*, 2001).

Patterns of plant genome evolution

Comparative genetic mapping has generally revealed collinear chromosomal segments in closely related plants, whereas comparative genome studies at the micro level have disclosed many small differences between genomes of closely related species. Even if genome segments of 100 kb or less are analysed deviations from collinearity are often apparent. Evidence for translocations and inversions, that involve one or several genes are readily detected (Figure 2). Particularly common, however, are deletions and duplications of gene sequences (Figures 1 and 2), possibly resulting from unequal crossing-over. All these results taken together indicate that alterations of the fine structure may play a much more prominent role in the evolution of plant genomes than gross chromosomal rearrangements.

In this context, it is interesting to note that the analysis of the *Arabidopsis* genome revealed 1528 tandem arrays containing 4140 individual genes. Moreover, it is striking that the proportion of proteins belonging to families of more than five members is much higher in *Arabidopsis* than the values that have been reported for *Drosophila* or *Caenorhabditis elegans* (Arabidopsis Genome Initiative, 2000). Taken together, the results of microsynteny studies and the analysis of the *Arabidopsis* genome sequence, indicate that tandem gene duplications may play an important role in shaping plant genomes.

Plasticity of genome microstructure is also seen if genomes of different ecotypes are compared. A region of *A. thaliana* chromosome 4, ecotype Columbia, was compared to the corresponding region of the Landsberg *erecta* accession. The region harbours two retroelement-like sequences in the Columbia ecotype,

whereas in Landsberg *erecta* three are found in different positions. Moreover, polymorphisms including both DNA sequence and copy number of genes in tandem arrays were observed (Noël *et al.*, 1999).

A comparative analysis of 82 Mb of *Arabidopsis* genome sequence, accession Columbia, and 92.1 Mb of non-redundant sequences of the Landsberg *erecta* ecotype detected 14 570 insertions/deletions which range in size from 2 bp to 38 kb. Insertions/deletions >250 bp in Columbia compared to Landsberg *erecta* genomic sequences are often caused by transposon insertion or excision; however, evidence for the translocation of genes to new locations in the genome is also frequently found (Arabidopsis Genome Initiative, 2000).

Non-collinear positioning of transposon sequences has been observed in many collinearity studies; moreover, the rapid divergence of these elements has been noted. The importance of retroelements in shaping plant genomes is particularly noteworthy in the large grass genomes (Bennetzen *et al.*, 1998). However, size differences in intergenic regions are not always explained by the presence of retroelement-like sequences. For example, intergenic regions in tomato are expanded in comparison to the orthologous segments in *Arabidopsis* and *Capsella*, but hallmarks of retrotransposons were not found (Rossberg *et al.*, 2001).

All synteny studies carried out to date between species belonging to the Brassicaceae show that *Arabidopsis* and *Capsella* display more pronounced conservation of genome structure than *Arabidopsis* and *Brassica* (Arabidopsis Genome Initiative, 2000; Bancroft, 2001; Schmidt *et al.*, 2001). The species *Arabidopsis* and *Capsella* diverged more recently than the lineages leading to *Arabidopsis* and *Brassica* (Acarkan *et al.*, 2000). Nevertheless, this does not fully account for the differences seen in comparative genome studies. In this context it is important to note that the most pronounced deviations from conservation in genome structure are seen in multiplied regions of the *Brassica* genome. The triplicated segments of the *Brassica* genomes differ in respect to gene repertoire and only the genes of the triplicated regions taken together make up the gene content in the corresponding *Arabidopsis* region. One or several homologues of *Arabidopsis* genes may be missing from any particular triplicated region. These results are consistent with the hypothesis that gene deletion events occur frequently in multiplied regions of a genome. The complex nature of the *Brassica* genome with many

regions being present in multiple copies may thus be the crucial factor for the less pronounced collinearity seen for *Arabidopsis* and *Brassica* when compared to *Arabidopsis* and *Capsella*.

Analysis of the *Arabidopsis* genome sequence data led to the discovery of large segmental duplications (Blanc *et al.*, 2000; Paterson *et al.*, 2000; Arabidopsis Genome Initiative, 2000; Vision *et al.*, 2000). In the segmental duplications seen in the *Arabidopsis* genome, a set of common genes is interspersed with genes unique to any one of the regions. Collinearity studies with distantly related species support the view that the segmental duplications indeed originate from a common ancestral chromosome segment (Ku *et al.*, 2000; Rossberg *et al.*, 2001). Therefore, the duplications in the *Arabidopsis* genome share the same characteristics as the multiplied segments in the *Brassica* genome. Studies of the *Arabidopsis* genome sequence thus also support the hypothesis that duplicated segments may substantially influence plant genome evolution.

In this context, a study of synthetic polyploids of *Brassica* by Song *et al.* (1995) is particularly noteworthy, since evidence for extensive and rapid genome change was presented. The described alterations could be due to different processes such as chromosomal rearrangements, point mutations or gene conversions. By contrast, the organization of the genomes of amphidiploid *B. napus* and *B. juncea* is very similar to the ones from their progenitors, thus polyploidization events are not necessarily followed by extensive alterations in chromosome structure (Axelsson *et al.*, 2000; Parkin *et al.*, 1995; Bohuon *et al.*, 1996).

The duplicated regions in the *Arabidopsis* genome may be remnants of single or multiple polyploidizations. Alternatively, they might represent independent segmental duplications. Two different analyses have shown that the majority of the genome falls into duplicated blocks. This was taken as a hint that the duplicated blocks are likely due to a single polyploidization event (Arabidopsis Genome Initiative, 2000). Consistent with that view, a molecular-clock analysis by Lynch and Conery (2000) identified a large group of duplicated genes which belong to the same age class. The age of these duplications has been estimated at 65 million years. In contrast, an independent study suggested at least four large-scale duplication events that occurred 100 to 200 million years ago (Vision *et al.*, 2000). However, as discussed by Wolfe (2001), analyses of phylogenetic trees and sequences from an outgroup are needed to confirm whether more than

one large-scale duplication event has occurred in the lineage leading to *Arabidopsis thaliana*.

Comparative mapping between *Arabidopsis*, *Capsella* and *Brassica* has provided evidence that at least the few *Arabidopsis* large-scale duplications studied predate the divergence of the three crucifer species (Bancroft, 2000; O'Neill and Bancroft, 2000; Arabidopsis Genome Initiative, 2000; Rossberg *et al.*, 2001). Thus, using more comparative data, especially with more distantly related species, will shed light on the age of the duplicated blocks.

It needs to be considered that polyploidy is widespread in the plant kingdom (Wendel, 2000). The analysis of the *Arabidopsis* genome has shown that large-scale duplications may even be discovered in plant species with very small genomes. Since large-scale duplications and subsequent gene loss seem to be a very important process in plant genome evolution it is of great significance to assess the occurrence of polyploidization events in different plant lineages. Only such studies will allow the use of comparative genomics in the most effective way.

Acknowledgement

I thank Dr B. Schulz (University of Tübingen, Germany) for helpful comments on the manuscript.

References

- Acarkan, A., Rossberg, M., Koch, M. and Schmidt, R. 2000. Comparative genome analysis reveals extensive conservation of genome organisation for *Arabidopsis thaliana* and *Capsella rubella*. *Plant J.* 23: 55–62.
- Ahn, S. and Tanksley, S.D. 1993. Comparative linkage maps of the rice and maize genomes. *Proc. Natl. Acad. Sci. USA* 90: 7980–7984.
- Arabidopsis Genome Initiative 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- Arumuganathan, K. and Earle, E.D. 1991. Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9: 208–218.
- Avramova, Z., Tikhonov, A., SanMiguel, P., Jin, Y.-K., Liu, C., Woo, S.-S., Wing, R.A. and Bennetzen, J.L. 1996. Gene identification in a complex chromosomal continuum by local genomic cross-referencing. *Plant J.* 10: 1163–1168.
- Axelsson, T., Bowman, C.M., Sharpe, A.G., Lydiate, D.J. and Lagercrantz, U. 2000. Amphidiploid *Brassica juncea* contains conserved progenitor genomes. *Genome* 43: 679–688.
- Bancroft, I. 2000. Insights into the structural and functional evolution of plant genomes afforded by the nucleotide sequences of chromosomes 2 and 4 of *Arabidopsis thaliana*. *Yeast* 17: 1–5.
- Bancroft, I. 2001. Duplicate and diverge: the evolution of plant genome microstructure. *Trends Genet.* 17: 89–93.

- Barry, G.F. 2001. The use of the Monsanto draft rice genome sequence in research. *Plant Physiol.* 125: 1164–1165.
- Bennett, M.D. 1998. Plant genome values: how much do we know? *Proc. Natl. Acad. Sci. USA* 95: 2011–2016.
- Bennett, M.D. and Smith, J.B. 1976. Nuclear DNA amounts in angiosperms. *Phil. Trans. R. Soc. Lond.* 274: 227–274.
- Bennett, M.D. and Smith, J.B. 1991. Nuclear DNA amounts in angiosperms. *Phil. Trans. R. Soc. Lond.* 334: 309–345.
- Bennett, M.D., Smith, J.B. and Heslop-Harrison, J.S. 1982. Nuclear DNA amounts in angiosperms. *Phil. Trans. R. Soc. Lond.* 216: 179–190.
- Bennetzen, J.L. 2000. Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. *Plant Cell* 12: 1021–1029.
- Bennetzen, J.L., SanMiguel, P., Chen, M., Tikhonov, A., Francki, M., Avramova, Z. 1998. Grass genomes. *Proc. Natl. Acad. Sci. USA* 95: 1975–1978.
- Blanc, G., Barakat, A., Guyot, R., Cooke, R. and Delseny, M. 2000. Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell* 12: 1093–1101.
- Bohuon, E.J.R., Keith, D.J., Parkin, I.A.P., Sharpe, A.G. and Lydiat, D.J. 1996. Alignment of the conserved C genomes of *Brassica oleracea* and *Brassica napus*. *Theor. Appl. Genet.* 93: 833–839.
- Bohuon, E.J.R., Ramsay, L.D., Craft, J.A., Arthur, A.E., Marshall, D.F., Lydiat, D.J. and Kearsley, M.J. 1998. The association of flowering time quantitative trait loci with duplicated regions and candidate loci in *Brassica oleracea*. *Genetics* 150: 393–401.
- Bonierbale, M.W., Plaisted, R.L. and Tanksley, S.D. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* 120: 1095–1103.
- Cavell, A.C., Lydiat, D.J., Parkin, I.A.P., Dean, C. and Trick, M. 1998. Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. *Genome* 41: 62–69.
- Chao, S., Sharp, P.J., Worland, A.J., Warham, E.J., Koebner, R.M.D. and Gale, M.D. 1989. RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. *Theor. Appl. Genet.* 78: 495–504.
- Chen, M., SanMiguel, P., de Oliveira, A.C., Woo, S.-S., Zhang, H., Wing, R.A. and Bennetzen, J.L. 1997. Microcolinearity in *sh2*-homologous regions of the maize, rice, and sorghum genomes. *Proc. Natl. Acad. Sci. USA* 94: 3431–3435.
- Chen, M., SanMiguel, P. and Bennetzen, J.L. 1998. Sequence organization and conservation in *sh2/a1*-homologous regions of sorghum and rice. *Genetics* 148: 435–443.
- Conner, J.A., Conner, P., Nasrallah, M.E. and Nasrallah, J.B. 1998. Comparative mapping of the *Brassica S* locus region and its homeolog in *Arabidopsis*: implications for the evolution of mating systems in the Brassicaceae. *Plant Cell* 10: 801–812.
- Devos, K.M. and Gale, M.D. 1993. The genetic maps of wheat and their potential in plant breeding. *Outl. Agric.* 22: 93–99.
- Devos, K.M. and Gale, M.D. 2000. Genome relationships: the grass model in current research. *Plant Cell* 12: 637–646.
- Devos, K.M., Beales, J., Nagamura, Y. and Sasaki, T. 1999. *Arabidopsis*-rice: will colinearity allow gene prediction across the eudicot-monocot divide? *Genome Res.* 9: 825–829.
- Devos, K.M., Pittaway, T.S., Reynolds, A. and Gale, M.D. 2000. Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. *Theor. Appl. Genet.* 100: 190–198.
- Draye, X., Lin, Y.R., Qian, X.-Y., Bowers, J.E., Burow, G.B., Morrell, P.L., Peterson, D.G., Presting, G.G., Ren, S.-X., Wing, R.A. and Paterson, A.H. 2001. Toward integration of comparative genetic, physical, diversity, and cytomechanical maps for grasses and grains, using the sorghum genome as a foundation. *Plant Physiol.* 125: 1325–1341.
- Druka, A., Kudrna, D., Han, F., Kilian, A., Steffenson, B., Frisch, D., Tomkins, J., Wing, R. and Kleinhofs, A. 2000. Physical mapping of the barley stem rust resistance gene *rpg4*. *Mol. Gen. Genet.* 264: 283–290.
- Dubcovsky, J., Luo, M.-C. and Dvorák, J. 1995. Differentiation between homoeologous chromosomes 1A of wheat and 1A^m of *Triticum monococcum* and its recognition by the wheat *Phl* locus. *Proc. Natl. Acad. Sci. USA* 92: 6645–6649.
- Dubcovsky, J., Ramakrishna, W., SanMiguel, P.J., Busso, C.S., Yan, L., Shiloff, B.A. and Bennetzen, J.L. 2001. Comparative sequence analysis of collinear barley and rice bacterial artificial chromosomes. *Plant Physiol.* 125: 1342–1353.
- Feuillet, C. and Keller, B. 1999. High gene density is conserved at syntenic loci of small and large grass genomes. *Proc. Natl. Acad. Sci. USA* 96: 8265–8270.
- Flavell, R. 1980. The molecular characterization and organization of plant chromosomal DNA sequences. *Annu. Rev. Plant Physiol.* 31: 569–596.
- Foot, T., Roberts, M., Kurata, N., Sasaki, T. and Moore, G. 1997. Detailed comparative mapping of cereal chromosome regions corresponding to the *Phl* locus in wheat. *Genetics* 147: 801–807.
- Fourmann, M., Barret, P., Renard, M., Pelletier, G., Delourme, R. and Brunel, D. 1998. The two genes homologous to *Arabidopsis FAE1* co-segregate with the two loci governing erucic acid content in *Brassica napus*. *Theor. Appl. Genet.* 96: 852–858.
- Fray, M.J., Puangsomlee, P., Goodrich, J., Coupland, G., Evans, E.J., Arthur, A.E. and Lydiat, D.J. 1997. The genetics of staminal petal production in oilseed rape (*Brassica napus*) and equivalent variation in *Arabidopsis thaliana*. *Theor. Appl. Genet.* 94: 731–736.
- Gale, M.D. and Devos, K.M. 1998. Comparative genetics in the grasses. *Proc. Natl. Acad. Sci. USA* 95: 1971–1974.
- Grant, M.R., McDowell, J.M., Sharpe, A.G., de Torres Zabala, M., Lydiat, D.J. and Dangl, J.L. 1998. Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 95: 15843–15848.
- Grant, D., Cregan, P. and Shoemaker, R.C. 2000. Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 97: 4168–4173.
- Han, F., Kilian, A., Chen, J.P., Kudrna, D., Steffenson, B., Yamamoto, K., Matsumoto, T., Sasaki, T. and Kleinhofs, A. 1999. Sequence analysis of a rice BAC covering the syntenous barley *Rpg1* region. *Genome* 42: 1071–1076.
- Helentjaris, T., Weber, D. and Wright, S. 1988. Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. *Genetics* 118: 353–363.
- Hulbert, S.H., Richter, T.E., Axtell, J.D. and Bennetzen, J.L. 1990. Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. *Proc. Natl. Acad. Sci. USA* 87: 4251–4255.
- Jackson, S.A., Cheng, Z.K., Wang, M.L., Goodman, H.M. and Jiang, J.M. 2000. Comparative fluorescence *in situ* hybridization mapping of a 431-kb *Arabidopsis thaliana* bacterial artificial chromosome contig reveals the role of chromosomal duplications in the expansion of the *Brassica rapa* genome. *Genetics* 156: 833–838.
- Kilian, A., Kudrna, D.A., Kleinhofs, A., Yano, M., Kurata, N., Steffenson, B. and Sasaki, T. 1995. Rice-barley synteny and

- its application to saturation mapping of the barley *Rpg1* region. *Nucl. Acids Res.* 23: 2729–2733.
- Kilian, A., Chen, J., Han, F., Steffenson, B. and Kleinhofs, A. 1997. Towards map-based cloning of the barley stem rust resistance genes *Rpg1* and *rpg4* using rice as an intergenomic cloning vehicle. *Plant Mol. Biol.* 35: 187–195.
- Kole, C., Quijada, P., Michaels, S.D., Amasino, R.M. and Osborn, T.C. 2001. Evidence for homology of flowering-time genes *VFR2* from *Brassica rapa* and *FLC* from *Arabidopsis thaliana*. *Theor. Appl. Genet.* 102: 425–430.
- Kowalski, S.P., Lan, T.-H., Feldmann, K.A. and Paterson, A.H. 1994. Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. *Genetics* 138: 499–510.
- Ku, H.-M., Vision, T., Liu, J. and Tanksley, S.D. 2000. Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. *Proc. Natl. Acad. Sci. USA* 97: 9121–9126.
- 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. and Lydiat, D. 1996. Comparative genome mapping in *Brassica*. *Genetics* 144: 1903–1910.
- Lagercrantz, U., Putterill, J., Coupland, G. and Lydiat, D. 1996. Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. *Plant J.* 9: 13–20.
- Lan, T.-H. and Paterson, A.H. 2000. Comparative mapping of quantitative trait loci sculpting the curd of *Brassica oleracea*. *Genetics* 155: 1927–1954.
- Lan, T.H., DelMonte, T.A., Reischmann, K.P., Hyman, J., Kowalski, S.P., McFerson, J., Kresovich, S., Paterson, A.H. 2000. An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. *Genome Res.* 10: 776–788.
- Leister, D., Kurth, J., Laurie, D.A., Yano, M., Sasaki, T., Devos, K., Graner, A. and Schulze-Lefert, P. 1998. Rapid reorganization of resistance gene homologues in cereal genomes. *Proc. Natl. Acad. Sci. USA* 95: 370–375.
- Leitch I.J., Chase M.W. and Bennett, M.D. 1998. Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants. *Ann. Bot.* 82 (Suppl. A): 85–94.
- Lin, Y.-R., Schertz, K.F. and Paterson, A.H. 1995. Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. *Genetics* 141: 391–411.
- Lin, X.Y., Kaul, S.S., Rounsley, S., Shea, T.P., Benito, M.-I., Town, C.D., Fujii, C.Y., Mason, T., Bowman, C.L., Barnstead, M., Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J., Ronning, C.M., Koo, H.L., Moffat, K.S., Cronin, L.A., Shen, M., Pai, G., Van Aken, S., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D., Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C. 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402: 761–768.
- Livingstone, K.D., Lackney, V.K., Blauth, J.R., van Wijk, R., Jahn, M.K. 1999. Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. *Genetics* 152: 1183–1202.
- Lynch, M. and Conery, J.S. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151–1155.
- Mayer, K., Schüller, C., Wambutt, R., Murphy, G., Volckaert, G., Pohl, T., Düsterhöft, A., Stiekema, W., Entian, K.-D., Terry, N., Harris, B., Ansoorge, W., Brandt, P., Grivell, L., Rieger, M., Weichselgartner, M., de Simone, V., Obermaier, B., Mache, R., Müller, M., Kreis, M., Delseny, M., Puigdomenech, P., Watson, M., Schmidheini, T., Reichert, B., Portatelle, D., Perez-Alonso, M., Boutry, M., Bancroft, I., Vos, P., Hoheisel, J., Zimmermann, W., Wedler, H., Ridley, P., Langham, S.-A., McCullagh, B., Bilham, L., Robben, J., Van der Schueren, J., Grymonprez, B., Chuang, Y.-J., Vandenbussche, F., Braeken, M., Weltjens, I., Voet, M., Bastiaens, I., Aert, R., Defoor, E., Weitzenegger, T., Bothe, G., Ramsperger, U., Hilbert, H., Braun, M., Holzer, E., Brandt, A., Peters, S., van Staveren, M., Dirkse, W., Mooijman, P., Klein Lankhorst, R., Rose, M., Hauf, J., Kötter, P., Bermeiser, S., Hempel, S., Feldpausch, M., Lamberth, S., Van den Daele, H., De Keyser, A., Buysschaert, C., Gielen, J., Villarreal, R., De Clercq, R., Van Montagu, M., Rogers, J., Cronin, A., Quail, M., Bray-Allen, S., Clark, L., Doggett, J., Hall, S., Kay, M., Lennard, N., McLay, K., Mayes, R., Pettett, A., Rajandream, M.-A., Lyne, M., Benes, V., Rechmann, S., Borkova, D., Blöcker, H., Scharfe, M., Grimm, M., Löhnert, T.-H., Dose, S., de Haan, M., Maarse, A., Schäfer, M., Müller-Auer, S., Gabel, C., Fuchs, M., Fartmann, B., Granderath, K., Dauner, D., Herzl, A., Neumann, S., Argiriou, A., Vitale, D., Liguori, R., Piravandi, E., Massenot, O., Quigley, F., Clabaud, G., Mündlein, A., Felber, R., Schnabl, S., Hiller, R., Schmidt, W., Lecharny, A., Aubourg, S., Chefedor, F., Cooke, R., Berger, C., Montfort, A., Casacuberta, E., Gibbons, T., Weber, N., Vandenbol, M., Barges, M., Terol, J., Torres, A., Perez-Perez, A., Purnelle, B., Bent, E., Johnson, S., Tacon, D., Jesse, T., Heijnen, L., Schwarz, S., Scholler, P., Heber, S., Francs, P., Bielke, C., Frishman, D., Haase, D., Lemcke, K., Mewes, H.W., Stocker, S., Zaccaria, P., Bevan, M., Wilson, R.K., de la Bastide, M., Habermann, K., Parnell, L., Dedhia, N., Gnoj, L., Schutz, K., Huang, E., Spiegel, L., Sehkon, M., Murray, J., Sheet, P., Cordes, M., Abu-Threideh, J., Stoneking, T., Kalicki, J., Graves, T., Harmon, G., Edwards, J., Latreille, P., Courtney, L., Cloud, J., Abbott, A., Scott, K., Johnson, D., Minx, P., Bentley, D., Fulton, B., Miller, N., Greco, T., Kemp, K., Kramer, J., Fulton, L., Mardis, E., Dante, M., Pepin, K., Hillier, L., Nelson, J., Spieth, J., Ryan, E., Andrews, S., Geisel, C., Layman, D., Du, H., Ali, J., Berghoff, A., Jones, K., Drone, K., Cotton, M., Joshu, C., Antoniou, B., Zidanic, M., Strong, C., Sun, H., Lamar, B., Yordan, C., Ma, P., Zhong, J., Preston, R., Vil, D., Shekher, M., Matero, A., Shah, R., Swaby, I.K., O’Shaughnessy, A., Rodriguez, M., Hoffman, J., Till, S., Granat, S., Shohdy, N., Hasegawa, A., Hameed, A., Lodhi, M., Johnson, A., Chen, E., Marra, M., Martienssen, R. and McCombie, W.R. 1999. Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* 402: 769–777.
- Messing, J. and Llaca, V. 1998. Importance of anchor genomes for any plant genome project. *Proc. Natl. Acad. Sci. USA* 95: 2017–2020.
- Moore, G., Foote, T., Helentjaris, T., Devos, K., Kurata, N. and Gale, M. 1995. Was there a single ancestral cereal chromosome? *Trends Genet.* 11: 81–82.
- Moore, G., Roberts, M., Aragon-Alcaide, L. and Foote, T. 1997. Centromeric sites and cereal chromosome evolution. *Chromosoma* 105: 321–323.
- Noël, L., Moores, T.L., van der Biezen, E.A., Parniske, M., Daniels, M.J., Parker, J.E. and Jones, J.D.G. 1999. Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of *Arabidopsis*. *Plant Cell* 11: 2099–2111.
- O’Neill, C.M. and Bancroft, I. 2000. Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homoeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. *Plant J.* 23: 233–243.

- Osborn, T.C., Kole, C., Parkin, I.A.P., Sharpe, A.G., Kuiper, M., Lydiate, D.J. and Trick, M. 1997. Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. *Genetics* 146: 1123–1129.
- Parkin, I.A.P., Sharpe, A.G., Keith, D.J. and Lydiate, D.J. 1995. Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). *Genome* 38: 1122–1131.
- Paterson, A.H., Lin, Y.-R., Li, Z., Schertz, K.F., Doebley, J.F., Pinson, S.R.M., Liu, S.-C., Stansel, J.W. and Irvine, J.E. 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* 269: 1714–1718.
- Paterson, A.H., Lan, T.-H., Reischmann, K.P., Chang, C., Lin, Y.-R., Liu, S.-C., Burow, M.D., Kowalski, S.P., Katsar, C.S., DelMonte, T.A., Feldmann, K.A., Schertz, K.F. and Wendel, J.F. 1996. Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. *Nature Genet.* 14: 380–382.
- Paterson, A.H., Bowers, J.E., Burow, M.D., Draye, X., Elisik, C.G., Jiang, C.-X., Katsar, C.S., Lan, T.-H., Lin, Y.-R., Ming, R. and Wright, R.J. 2000. Comparative genomics of plant chromosomes. *Plant Cell* 12: 1523–1539.
- Pavy, N., Rombauts, S., Dehais, P., Mathe, C., Ramana, D.V., Leroy, P., Rouzé, P. 1999. Evaluation of gene prediction software using a genomic data set: application to *Arabidopsis thaliana* sequences. *Bioinformatics* 15: 887–899.
- Peng, J., Richards, D.E., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintham, J.E., Beales, J., Fish, L.J., Worland, A.J., Pelica, F., Sudhakar, D., Christou, P., Snape, J.W., Gale, M.D. and Harberd, N.P. 1999. 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* 400: 256–261.
- Quiros, C.F., Grellet, F., Sadowski, J., Suzuki, T., Li, G. and Wroblewski, T. 2001. *Arabidopsis* and *Brassica* comparative genomics: sequence, structure and gene content in the *ABI1-Rps2-Ck1* chromosomal segment and related regions. *Genetics* 157: 1321–1330.
- Rabinowicz, P.D., Schutz, K., Dedhia, N., Yordan, C., Parnell, L.D., Stein, L., McCombie, W.R. and Martienssen, R.A. 1999. Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. *Nature Genet.* 23: 305–308.
- Rossberg, M., Theres, K., Acarkan, A., Herrero, R., Schmitt, T., Schumacher, K., Schmitz, G. and Schmidt, R. 2001. Comparative sequence analysis reveals extensive microcolinearity in the *Lateral suppressor* regions of the tomato, *Arabidopsis* and *Capsella* genomes. *Plant Cell* 13: 979–988.
- Sadowski, J. and Quiros, C.F. 1998. Organization of an *Arabidopsis thaliana* gene cluster on chromosome 4 including the *RPS2* gene in the *Brassica nigra* genome. *Theor. Appl. Genet.* 96: 468–474.
- Sadowski, J., Gaubier, P., Delseny, M. and Quiros, C.F. 1996. Genetic and physical mapping in *Brassica* diploid species of a gene cluster defined in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 251: 298–306.
- SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z. and Bennetzen, J.L. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274: 765–768.
- Scheffler, J.A., Sharpe, A.G., Schmidt, H., Sperling, P., Parkin, I.A.P., Lühs, W., Lydiate, D.J. and Heinz, E. 1997. Desaturase multigene families of *Brassica napus* arose through genome duplication. *Theor. Appl. Genet.* 94: 583–591.
- Schmidt, R. 2000. Synteny: recent advances and future prospects. *Curr. Opin. Plant Biol.* 3: 97–102.
- Schmidt, R., Acarkan, A. and Boivin, K. 2001. Comparative structural genomics in the Brassicaceae family. *Plant Phys. Biochem.* 39: 253–262.
- Sillito, D., Parkin, I.A.P., Mayerhofer, R., Lydiate, D.J. and Good, A.G. 2000. *Arabidopsis thaliana*: a source of candidate disease-resistance genes for *Brassica napus*. *Genome* 43: 452–460.
- Slocum, M.K., Figdore, S.S., Kennard, W.C., Suzuki, J.Y. and Osborn, T.C. 1990. Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. *Theor. Appl. Genet.* 80: 57–64.
- Soltis, P.S., Soltis, D.E. and Chase, M.W. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402: 402–404.
- Song, K.M., Suzuki, J.Y., Slocum, M.K., Williams, P.H. and Osborn, T.C. 1991. A linkage map of *Brassica rapa* (syn. *campestris*) based on restriction fragment length polymorphism loci. *Theor. Appl. Genet.* 82: 296–304.
- Song, K.M., Lu, P., Tang, K.L. and Osborn, T.C. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proc. Natl. Acad. Sci. USA* 92: 7719–7723.
- Stein, N., Feuillet, C., Wicker, T., Schlagenhauf, E. and Keller, B. 2000. Subgenome chromosome walking in wheat: a 450-kb physical contig in *Triticum monococcum* L. spans the *Lr10* resistance locus in hexaploid wheat (*Triticum aestivum* L.). *Proc. Natl. Acad. Sci. USA* 97: 13436–13441.
- Tanksley, S.D., Bernatzky, R., Lapitan, N.L. and Prince, J.P. 1988. Conservation of gene repertoire but not gene order in pepper and tomato. *Proc. Natl. Acad. Sci. USA* 85: 6419–6423.
- Tanksley, S.D., Ganai, M.W., Prince, J.P., de Vicente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., Messeguer, R., Miller, J.C., Miller, L., Paterson, A.H., Pineda, O., Röder, M.S., Wing, R.A., Wu, W. and Young, N.D. 1992. High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132: 1141–1160.
- Tarchini, R., Biddle, P., Wineland, R., Tingey, S. and Rafalski, A. 2000. The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted colinearity with maize chromosome 4. *Plant Cell* 12: 381–391.
- Terryn, N., Heijnen, L., De Keyser, A., Van Asseldonck, M., De Clercq, R., Verbakel, H., Gielen, J., Zabeau, M., Villarroel, R., Jesse, T., Neyt, P., Hogers, R., Van Den Daele, H., Ardiles, W., Schueller, C., Mayer, K., Déhais, P., Rombauts, S., Van Montagu, M., Rouzé, P. and Vos, P. 1999. Evidence for an ancient chromosomal duplication in *Arabidopsis thaliana* by sequencing and analyzing a 400-kb contig at the *Apetala2* locus on chromosome 4. *FEBS Lett.* 445: 237–245.
- Tikhonov, A.P., SanMiguel, P.J., Nakajima, Y., Gorenstein, N.M., Bennetzen, J.L. and Avramova, Z. 1999. Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. *Proc. Natl. Acad. Sci. USA* 96: 7409–7414.
- Truco, M.J. and Quiros, C.F. 1994. Structure and organization of the B genome based on a linkage map in *Brassica nigra*. *Theor. Appl. Genet.* 89: 590–598.
- van Dodeweerd, A.M., Hall, C.R., Bent, E.G., Johnson, S.J., Bevan, M.W. and Bancroft, I. 1999. Identification and analysis of homoeologous segments of the genomes of rice and *Arabidopsis thaliana*. *Genome* 42: 887–892.
- Vision, T.J., Brown, D.G. and Tanksley, S.D. 2000. The origins of genomic duplications in *Arabidopsis*. *Science* 290: 2114–2117.
- Wendel, J.F. 2000. Genome evolution in polyploids. *Plant Mol. Biol.* 42: 225–249.
- Wolfe, K.H. 2001. Yesterday's polyploids and the mystery of diploidization. *Nat. Rev. Genet.* 2: 333–341.

- Wu, J., Kurata, N., Tanoue, H., Shimokawa, T., Umehara, Y., Yano, M. and Sasaki, T. 1998. Physical mapping of duplicated genomic regions of two chromosome ends in rice. *Genetics* 150: 1595–1603.
- Yuan, Q., Quackenbush, J., Sultana, R., Perte, M., Salzberg, S.L. and Buell, C.R. 2001. Rice bioinformatics. Analysis of rice sequence data and leveraging the data to other plant species. *Plant Physiol.* 125: 1166–1174.