



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
Main Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2002

Comparative Genomics in the Grass Family: Molecular Characterization of Grass Genome Structure and Evolution

Feuillet, C

Abstract: The genomes of grasses are very different in terms of size, ploidy level and chromosome number. Despite these significant differences, it was found by comparative mapping that the linear order (colinearity) of genetic markers and genes is very well conserved between different grass genomes. The potential of such conservation has been exploited in several directions, e.g. in defining rice as a model genome for grasses and in designing better strategies for positional cloning in large genomes. Recently, the development of large insert libraries in species such as maize, rice, barley and diploid wheat has allowed the study of large stretches of DNA sequence and has provided insight into gene organization in grasses. It was found that genes are not distributed randomly along the chromosomes and that there are clusters of high gene density in species with large genomes. Comparative analysis performed at the DNA sequence level has demonstrated that colinearity between the grass genomes is retained at the molecular level (microcolinearity) in most cases. However, detailed analysis has also revealed a number of exceptions to microcolinearity, which have given insight into mechanisms that are involved in grass-genome evolution. In some cases, the use of rice as a model to support gene isolation from other grass genomes will be complicated by local rearrangements. In this Botanical Briefing, we present recent progress and future prospects of comparative genomics in grasses

DOI: <https://doi.org/10.1093/aob/mcf008>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-154795>

Journal Article

Published Version

Originally published at:

Feuillet, C (2002). Comparative Genomics in the Grass Family: Molecular Characterization of Grass Genome Structure and Evolution. *Annals of Botany*, 89(1):3-10.

DOI: <https://doi.org/10.1093/aob/mcf008>

BOTANICAL BRIEFING

Comparative Genomics in the Grass Family: Molecular Characterization of Grass Genome Structure and Evolution

CATHERINE FEUILLET* and BEAT KELLER

Institute of Plant Biology, University of Zürich, Zollikerstrasse 107, CH-8008 Zürich, Switzerland

Received: 3 July 2001 Returned for revision: 4 September 2001 Accepted: 20 September 2001

The genomes of grasses are very different in terms of size, ploidy level and chromosome number. Despite these significant differences, it was found by comparative mapping that the linear order (colinearity) of genetic markers and genes is very well conserved between different grass genomes. The potential of such conservation has been exploited in several directions, e.g. in defining rice as a model genome for grasses and in designing better strategies for positional cloning in large genomes. Recently, the development of large insert libraries in species such as maize, rice, barley and diploid wheat has allowed the study of large stretches of DNA sequence and has provided insight into gene organization in grasses. It was found that genes are not distributed randomly along the chromosomes and that there are clusters of high gene density in species with large genomes. Comparative analysis performed at the DNA sequence level has demonstrated that colinearity between the grass genomes is retained at the molecular level (microcolinearity) in most cases. However, detailed analysis has also revealed a number of exceptions to microcolinearity, which have given insight into mechanisms that are involved in grass-genome evolution. In some cases, the use of rice as a model to support gene isolation from other grass genomes will be complicated by local rearrangements. In this Botanical Briefing, we present recent progress and future prospects of comparative genomics in grasses.

© 2002 Annals of Botany Company

Key words: Review, colinearity, gene density, genome evolution, genome structure, grasses (Poaceae), microcolinearity.

INTRODUCTION

In the last 10 years, comparative mapping in plants has provided evidence for a remarkable conservation of marker and gene order (colinearity) between related genomes and has resulted in the new discipline of comparative genomics. Comparative studies performed on grasses have provided the most comprehensive data set to date on colinearity between genomes within a large plant family. Most of these analyses have been performed on economically important grass species such as the staple cereals, rice, wheat, barley, maize, millet, oat and sorghum (for reviews, see Gale and Devos, 1998; Bennetzen, 2000a; Devos and Gale, 2000; Keller and Feuillet, 2000). In maize, rice, sorghum, barley and wheat, comparisons have been extended to the DNA sequence level (microcolinearity), allowing study of the conservation of coding and non-coding regions as well as characterization of molecular mechanisms of genome evolution in the grasses. As knowledge about the extent of conservation between the grass genomes has accumulated, the potential and the limits of the use of colinearity have become clearer. The conservation of genomes has been exploited to define sets of well-conserved anchor-probes (Van Deynze *et al.*, 1998), which are particularly useful when establishing genetic maps in grass species that are not well studied so far. The use of molecular markers derived from orthologous regions in different grass species has helped to increase the map density at specific genetic loci

and facilitates map-based cloning (Kilian *et al.*, 1997). Finally, comparative studies have allowed rice to be promoted as the model genome for grasses, leading to large projects such as the rice genome sequencing initiative (<http://rgp.dna.affrc.go.jp/Seqcollab.html>) that should result in an improved understanding of grass genomes in general. A good knowledge of genome organization is also necessary to define the best strategies and the tools necessary to isolate genes of agronomic importance from large and complex cereal genomes. Microcolinearity studies provide key information on the genome structure and the mechanisms responsible for differences in genome size and evolution in grasses.

After briefly introducing the main achievements of the last 10 years in the field of comparative genetics in grasses, we focus on comparative analysis at the DNA level (microcolinearity) and its impact on our understanding of genome organization and genome evolution in grasses.

COMPARATIVE ANALYSIS IN GRASSES: LEADING THE FIELD OF PLANT COMPARATIVE GENETICS

Comparative mapping: evidence for a remarkable colinearity between the grass genomes

The first preliminary studies concerning comparative mapping in plants were performed on the Solanaceae with the demonstration that cDNA markers were colinear along the tomato and potato chromosomes

*For correspondence: Fax 00 4116348201, e-mail feuillet@botinst.unizh.ch

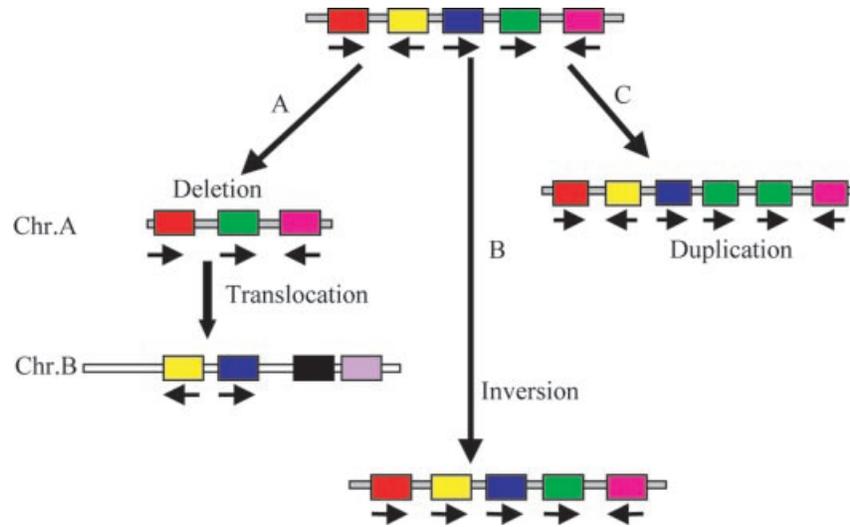


FIG. 1. Different types of micro-rearrangements observed in grass genomes at the microcolinearity level: deletion and/or translocation of small DNA fragments to another chromosome (A) (Kilian *et al.*, 1997; Tikhonov *et al.*, 1999; Tarchini *et al.*, 2000); gene inversion (B) or gene duplication (C) (Chen *et al.*, 1997; Dubcovsky *et al.*, 2001) or a combination of these rearrangements. Genetic mapping using probes corresponding to the adjacent sequences would indicate colinearity in this region. However, some of these micro-rearrangements (e.g. translocations) will complicate the analysis and limit the use of colinearity in cross genome map-based cloning strategies. Different genes are indicated by coloured boxes and the orientation of the transcription is indicated by an arrow below the genes.

(Bonierbale *et al.*, 1988). Soon after, the first comparative mapping studies were performed on grasses. These studies revealed a high degree of conservation of the map position and order (colinearity) of many markers between chromosomal regions of different grass genomes. Such colinearity was remarkable given differences in genome size of up to 40-fold and evolutionary divergence times of more than 60 000 000 years between the grass species (see Devos and Gale, 1997; Gale and Devos, 1998; Keller and Feuillet, 2000). Moreover, quantitative trait loci (QTL) underlying important agronomic traits such as shattering and dwarfing were also found to be colinear between grass species (Paterson *et al.*, 1995; Pereira and Lee, 1995; Wang *et al.*, 2001). The first grass consensus map aligning the genomes of seven different grass species using rice as a reference genome was published in 1995 (Moore *et al.*, 1995) and is regularly refined (Gale and Devos, 1998; Devos and Gale, 2000). Ten grass genomes can be described using less than 30 rice linkage blocks (Devos and Gale, 2000), and comparative mapping in grasses has resulted in the most comprehensive data set of comparative genomics in a plant family to date.

Does colinearity reflect microcolinearity?

The remarkable conservation of the marker order at the genetic map level raised the question of whether colinearity is retained at the molecular level. The first comparative studies of gene organization at the molecular level were performed between small genomic regions corresponding to two maize loci (*sh2/a1* and *Adh1*) and the homologous

regions in sorghum and rice. Restriction mapping and partial sequencing at the *sh2/a1* loci demonstrated that gene order and composition was conserved between maize, sorghum and rice (Chen *et al.*, 1997). Microcolinearity was also found between the large wheat and barley genomes at loci encoding orthologous receptor-like kinases (Feuillet and Keller, 1999). These first data suggested that microcolinearity is well retained between the grass genomes. However, a number of studies revealed significant gene rearrangements within otherwise microcolinear regions (Fig. 1). At the *Adh1* locus, sequence comparisons between maize and sorghum (Tikhonov *et al.*, 1999) and the more distantly related rice (Tarchini *et al.*, 2000) have shown that deletions/insertions or translocations of genes have occurred during evolution. A similar local lack of microcolinearity was also observed between the stem rust resistance gene *rpg1* locus in barley and the orthologous region in rice. In this case, the relocation of a small 10–15 kb DNA fragment was responsible for the lack of microcolinearity in the vicinity of the resistance gene (Kilian *et al.*, 1997). Thus, within microcolinear regions, different types of small rearrangements are likely to occur without rearrangement of the adjacent sequences (Fig. 1). Consequently, these regions will appear as colinear at the genetic map level although micro-rearrangements have occurred. Clearly, some rearrangements such as small inversions or gene duplications (Fig. 1B, C) will have little effect on microcolinearity whereas deletions and translocations (Fig. 1A) can greatly complicate the analysis. Thus, the use of rice as a model for the map-based isolation of genes from other grass genomes may frequently be complicated by such local genome rearrangements. Consequently, approaches based

on colinearity between grass genomes must also be performed using more closely related species, e.g. within tribes or subtribes.

Genomic tools and strategies for the isolation of genes from large grass genomes

Recently, a number of initiatives have been undertaken to develop new tools such as large insert libraries, EST (expressed sequence tag) collections, physical maps and gene targeting systems in grass species with large genomes such as maize, barley and wheat (see <http://www.agron.missouri.edu/index.html>; <http://wheat.pw.usda.gov/NSF>). For example, in the last 2 years tremendous effort has been put into improving the availability of genomic tools for wheat. The two first wheat BAC (bacterial artificial chromosome) libraries were constructed in 1999 (Lijavetzky *et al.*, 1999; Moullet *et al.*, 1999) and an international initiative (ITEC: <http://wheat.pw.usda.gov/genome/>) was launched in 1998 to increase the number of ESTs from approx. five in 1999 to more than 58 000 in wheat and 68 000 in barley today. These efforts have already resulted in the successful development of new strategies for gene isolation from hexaploid wheat. Stein *et al.* (2000) have exploited the conservation between the homoeologous A genomes of the diploid einkorn wheat *Triticum monococcum* L. and the hexaploid wheat *Triticum aestivum* L. to perform 'subgenome' chromosome walking. In this approach, the colinearity between the *T. monococcum* and *T. aestivum* genomes was used for genetic mapping in hexaploid wheat and chromosome walking using the BAC library of *T. monococcum*. BAC clones were isolated after screening the *T. monococcum* library with an RFLP (restriction fragment length polymorphism) marker co-segregating with the leaf rust resistance gene *Lr10* in hexaploid wheat. Low-copy probes were then identified after low-pass sequencing of the BACs and were mapped in hexaploid wheat. Using this strategy, one step of subgenome chromosome walking was sufficient to establish a physical contig of approx. 300 kb in *T. monococcum*, which genetically spans the *Lr10* gene in hexaploid wheat (Stein *et al.*, 2000).

Even in the absence of local microcolinearity, the overall good colinearity observed between the grass genomes still offers the possibility of increasing the number of markers in a targeted region using RFLP and EST probes without the need to develop additional markers from the species of interest. This approach has already been used successfully to saturate different genomic regions of sugar cane, barley and wheat (Kilian *et al.*, 1997; Roberts *et al.*, 1999; Asnaghi *et al.*, 2000; Druka *et al.*, 2000).

Can colinearity be exploited across the monocot/dicot divide?

Several studies have been performed to define the extent to which colinearity is retained between monocotyledonous and dicotyledonous plants and to find out whether knowledge accumulating on *Arabidopsis thaliana* can be used to support gene isolation in grasses. An early predictive model

suggested that colinearity between sorghum and *A. thaliana* chromosomal segments could be expected within a distance < 3 cM (Paterson *et al.*, 1996). Sequence comparison of the *Adh1* and *Sh2/A1* regions of maize and sorghum with the *Arabidopsis* genome showed that adjacent genes in grasses are generally not colinear with *Arabidopsis* (Bennetzen *et al.*, 1998; Tikhonov *et al.*, 1999). Comparative analysis was also performed between *A. thaliana* and rice, the two model species for dicots and monocots, respectively. In one study, a conserved framework of genes (i.e. conserved colinearity for five genes interspersed by a number of non-conserved genes) was identified between both species in a region spanning 2.1 cM in rice (van Dodeweerd *et al.*, 1999). In contrast, Devos *et al.* (1999) did not find any supporting evidence that comparative mapping between rice and *A. thaliana* can be helpful in isolating genes from monocots. This lack of colinearity might be explained by the recently discovered repeated rounds of large-scale genome duplication and selective gene loss that have marked the evolution of *A. thaliana* (Vision *et al.*, 2000).

However, *A. thaliana* is useful for the isolation of genes that are involved in basic developmental processes and that are highly conserved between dicots and monocots. This has been demonstrated for the 'green revolution' dwarfing genes by Peng *et al.* (1999). In their study, the authors cloned cereal homologues of a gibberellic acid insensitive gene (*GAI*) involved in dwarfism in *A. thaliana* using rice EST sequence information. These homologues mapped to the same location as the *Rht-1* semi-dwarfing genes in wheat and the *d8* dwarf mutation in maize, showing the power of EST databases in establishing a link between *A. thaliana* genes and their homologues in cereals. Thus, although colinearity with *A. thaliana* will be very difficult to exploit for supporting gene isolation by map-based cloning from grasses, ESTs could help to identify grass genes based on the knowledge of their function in *A. thaliana*.

GENOME STRUCTURE AND MECHANISMS OF GENOME EVOLUTION IN GRASSES

Genome structure in large grass genomes

DNA sequence comparisons of large regions in different grass genomes have shown that coding regions are usually well conserved, but that the distances between the genes seem to be correlated with genome size (see Bennetzen, 2000a, and references therein; Fig. 2A). For example, intergenic regions are much larger in maize than in rice and sorghum at the *Adh1* and *sh2/a1* loci (Avramova *et al.*, 1996; Chen *et al.*, 1997, 1998) and are larger in barley than in rice at the orthologous *Fr-1/Hd6* regions (Dubcovsky *et al.*, 2001). Based on these observations, large intergenic distances could be expected in the wheat and barley genomes. However, regions of high-gene density have also been found in these genomes. In barley, a density of approx. one gene every 20 kb was observed in two regions of 60 kb and 66 kb at the *mlo* and *Rar1* loci (Panstruga *et al.*, 1998; Shirasu *et al.*, 2000). Moreover, high-gene density regions are conserved between the wheat and barley genomes (Fig. 2B). At the *Lrk10* locus in wheat and its

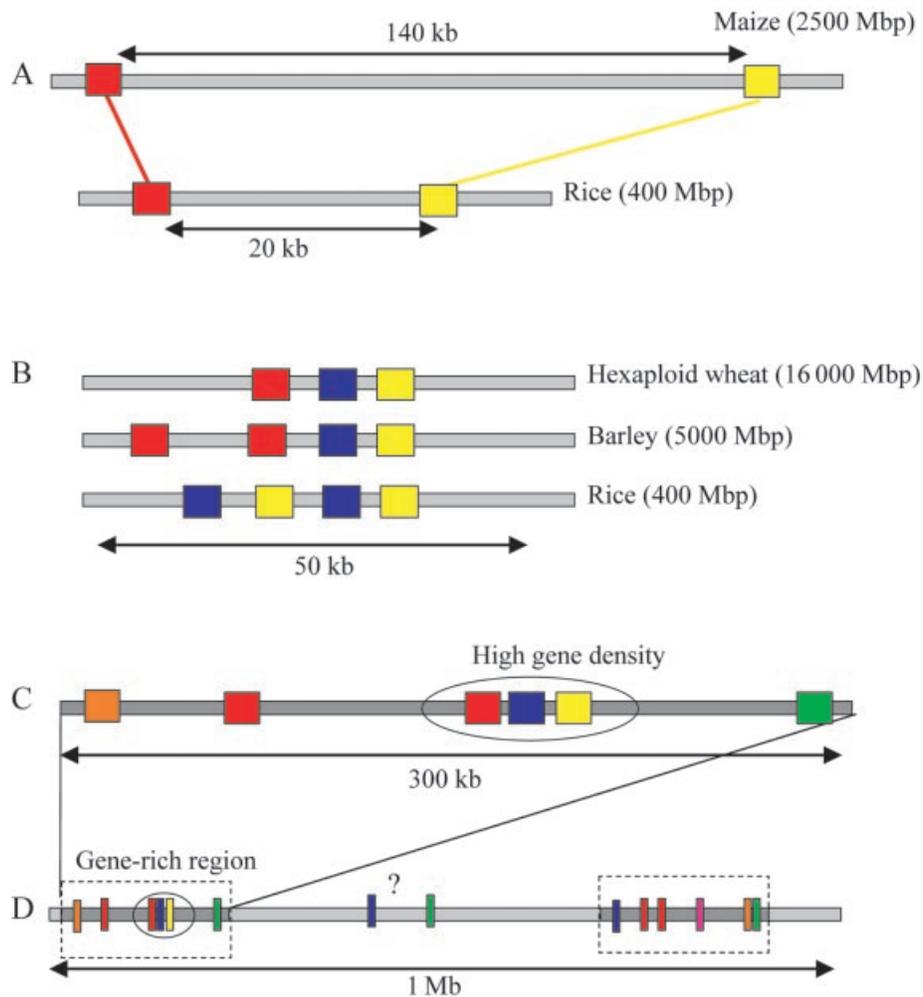


FIG. 2. Different types of gene organization observed in grass genomes. A, Coding regions are conserved between maize and rice but the size of the intergenic region varies according to the genome size (Chen *et al.*, 1997). B, Similar high-gene density regions are found at orthologous loci in genomes which size can differ by a factor > 12 (Feuillet *et al.*, 1999). C, Both high-gene density regions and genes interspersed by large intergenic regions coexist on the same DNA fragment (Tikhonov *et al.*, 1999; Wicker *et al.*, 2001). D, Model for the large scale gene organization and evolution of large grass genomes. Gene-rich regions, which are composed of both high-gene density islands (one gene every 5–20 kb) and single genes interspersed by less than 150 kb distances, are distributed along the chromosomes. The question mark indicates that the presence of genes in the large regions located in-between the gene-rich regions has not yet been demonstrated. Genes are represented by different coloured boxes.

orthologous region in barley, a gene density of one gene per 4–5 kb was observed, similar to that found in *A. thaliana* (Feuillet and Keller, 1999). The sequencing and comparison of larger DNA stretches refined the picture of gene organization in grasses. At the *Adh1* locus in maize, sequence analysis of 225 kb and comparison with 78 kb of the homologous region of sorghum revealed that, although large intergenic regions (70 kb) exist between genes, some genes are clustered within small distances (four genes within 39 kb of sequence without repetitive elements; Tikhonov *et al.*, 1999). Thus, within a distance of approx. 250 kb a combination of high gene density clusters was present beside individual genes separated from each other by large distances. Sequence analysis of 211 kb from *T. monococcum* has recently identified a similar pattern of gene distribution in wheat. The five genes which are present on the sequenced fragment are not distributed equally: three

genes are found clustered within 31 kb in a hot spot of recombination while the two others are separated from the cluster and from each other by large stretches (30–140 kb) of repetitive DNA (Wicker *et al.*, 2001).

So far, comparative studies have focused mainly on gene-containing regions. In species with large genomes such as wheat and barley, these regions have an overall gene density of one gene every 5–20 kb, which is higher than expected based on the random distribution of genes along the chromosomes (one gene every 250 kb: see Keller and Feuillet, 2000; Sandhu *et al.*, 2001). Thus, gene-containing regions seem to correspond to gene-rich regions that are made of single genes and high-gene density islands (Fig. 2C). In the near future, the sequencing and comparison of larger fragments (> 1 Mb) from gene-containing and also gene-poor regions will provide more information about grass genome structure in general. It will be interesting to

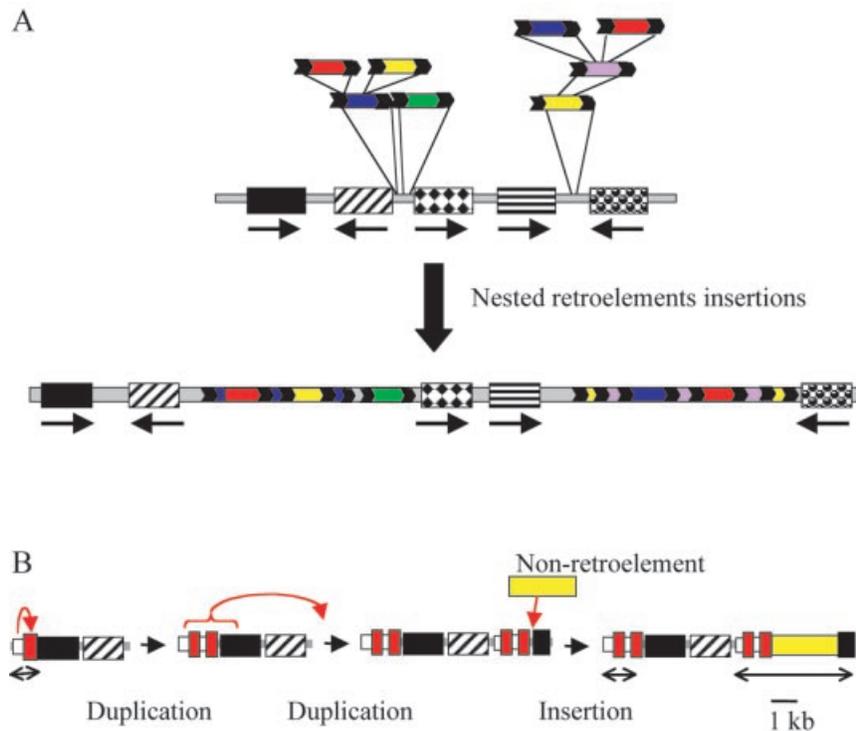


FIG. 3. Possible mechanisms of genome expansion in the grass genomes. A, Insertion of retroelements in the intergenic regions is a major driving force of genome expansion. Several waves of retroelement invasions can occur leading to the insertion of retrotransposons within each other (nested retroelements; SanMiguel *et al.*, 1996; Wicker *et al.*, 2001). B, Local duplications and insertion of sequences that do not show features of retroelements can also contribute to genome size increase (Feuillet *et al.*, 2001; Wicker *et al.*, 2001). Genes are indicated as black and white patterned boxes. Retrotransposons are represented by coloured chevrons flanked by black chevrons representing the long terminal repeats (LTRs).

see whether the genes are found exclusively within gene-rich regions or are also distributed in the gene-poor and more repetitive regions (Fig. 2D).

Comparative genomics as a tool to elucidate genome evolution in grasses

A very well-studied phylogeny (Kellogg, 1998, 2001) combined with comparative genomics make grasses the system of choice to study the mechanisms of plant genome evolution. Comparative studies indicate that grasses have undergone many events of genome expansion, contraction and rearrangements, but have maintained a remarkable overall conservation between the genomes (Kellogg, 1998; Gaut *et al.*, 2000). The study by Kellogg (1998) on the relationship between C-value (amount of DNA/diploid mitotic nucleus) and phylogeny indicated that the C-value has increased and decreased in the same lineage over evolutionary time in grass species. What mechanisms might be responsible for both genome expansion and DNA loss, and how are genes and intergenic regions affected by these changes?

Comparative analysis of large DNA sequences reveals the driving forces of genome rearrangements and expansion

All comparative studies have demonstrated a high degree of conservation of the exons between homologous grass

genes. Intron positions are also conserved but their size can sometimes differ. However, this size difference is not necessarily correlated with genome size (Dubcovsky *et al.*, 2001). In contrast, intergenic regions are generally not conserved between homologous regions in different grass species. They differ in size and in sequence and their expansion is regarded as the main factor for size differences between the grass genomes. The first microcolinearity studies in grasses have helped to demonstrate the major role played by transposable elements in determining these differences (for a review, see Bennetzen, 2000b). Sequence comparisons of large DNA stretches have shown that the size difference between intergenic regions in maize and sorghum was mainly due to the presence of repetitive DNA, most of it corresponding to nested retrotransposons (SanMiguel *et al.*, 1996; Chen *et al.*, 1997; Tikhonov *et al.*, 1999). Retrotransposons accounted for more than 74 % of the 225 kb sequence present at the *Adh1* locus in maize whereas the corresponding region in sorghum which is approx. 78 kb in length did not contain any retroelements (Tikhonov *et al.*, 1999). Similar elements are probably involved in the expansion of intergenic regions in wheat and barley genomes. Dubcovsky *et al.* (2001) found large size differences between two intergenic regions at orthologous loci in rice and barley: 1.4 kb and 0.6 kb in rice corresponded to 24 kb and 30.5 kb in barley, respectively. Eighty per cent of the size difference could be explained by the insertion of different types of

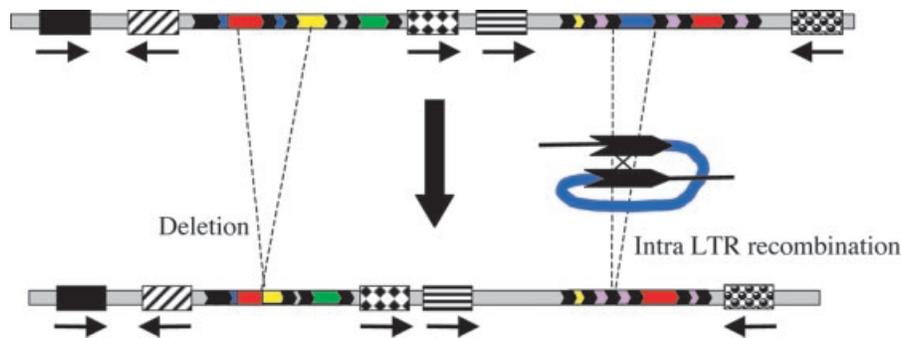


FIG. 4. Possible mechanisms leading to genome contraction. Genome contraction can be due to unequal crossing-over or intra-element recombination between nearby long terminal repeats (LTRs). Such recombination leads to the removal of the internal part of the retroelement (in blue) leaving a solo LTR (Vicient *et al.*, 1999; Shirasu *et al.*, 2000). Deletions of large DNA fragments consisting of different types of retroelements involve a mechanism independent from retrotransposon activity and are also responsible for DNA loss (Wicker *et al.*, 2001). Genes are indicated by black and white patterned boxes, retroelements are indicated by coloured chevrons delimited by LTRs represented as black chevrons.

retroelements in barley. The analysis of a 211 kb sequence from *T. monococcum* also showed that large intergenic regions are mainly composed of retroelements in wheat (Wicker *et al.*, 2001). In this case, 70 % of the sequence was composed of repetitive elements including ten types of retrotransposons, which showed a similar pattern of nested insertions as found in maize (SanMiguel *et al.*, 1996) and barley (Shirasu *et al.*, 2000). These studies suggest that retroelements play a major role in shaping and remodelling the genomes during evolution (Fig. 3A). A remarkable finding was that the maize genome size has doubled after an invasion of retrotransposons within the last 3000 000 years, after maize and sorghum diverged (SanMiguel *et al.*, 1998). This indicates that activity of retroelements can vary between closely related lineages providing a possible explanation for the C-value paradox.

A number of retroelements have been identified in different grass genomes (see Kumar and Bennetzen, 1999). The first studies on the conservation of these elements between the genomes suggested that retroelements are specific for species of the same genus or tribe (see Bennetzen, 1996). In a recent study, Vicient *et al.* (2001) searched the EST databases for expressed homologues of 38 known plant retrotransposons. They found that retrotransposons are generally more transcriptionally active and more conserved in grasses than in dicots. They also demonstrated that retroelements of the BARE-1 type (the most abundant retrotransposon family in barley) are present in different subfamilies of grasses outside the Triticeae tribe. In addition, polymorphisms were detected between different cultivars of wheat, oat, timothy and cordgrass suggesting that BARE-1-like elements have been active within these groups since their divergence from their last common ancestor. Comparison of the orthologous gene sequences in diploid and hexaploid wheat has shown the presence of a copia-like retrotransposon in the promoter region of the *T. monococcum* gene but not in the *T. aestivum* orthologue (unpubl. res.). This suggests that some retroelements have also been active since the divergence of the different wheat lineages. Langdon *et al.* (2000) have recently shown that a single ancestral family of retro-

transposons related to the Ty3-gypsy family is the source of all Poaceae centromere-specific retroelement sequences described to date. Further comparative studies on the evolution of transposable elements in relation to phylogeny improve understanding of their evolution and their impact on the grass genomes.

From the first sequence analysis in grass genomes, we can estimate that approx. 70 % of the intergenic regions are composed of retroelements. So far, few studies have focused on the type, the origin and the mechanisms driving the evolution of the remaining part of intergenic sequences. In wheat, a detailed analysis of 211 kb of sequence has shown that 30 % of the intergenic sequence consists of non-retrotransposon elements. Among them, seven foldback elements and three types of elements that do not show features of retroelements have been identified (Wicker *et al.*, 2001).

Comparative analysis can also help to discover new mechanisms that contribute to intergenic sequence evolution. Paralogous genomic regions, which have recently diverged, are ideal sequences in which to study such mechanisms. A striking example was recently reported in wheat where paralogous receptor-like kinase loci located on chromosome 1B were compared at the DNA sequence level (Feuillet *et al.*, 2001). A detailed comparison of the sequences flanking the conserved coding regions (> 95% identity) identified local duplications, insertions and deletions. The pattern allowed the establishment of a putative chronology for small and local rearrangements that have occurred before and after the duplication of the locus. These events did not involve repetitive elements but led to a size increase of 5–8 kb in the two paralogous regions (Fig. 3B). Further comparisons of recently duplicated regions will certainly provide more information about the mechanisms that participate in local genome expansion.

Which mechanisms counteract genome expansion?

Since the mechanisms responsible for genome expansion started to be unravelled, questions have arisen about the existence of counteracting mechanisms to avoid genome

explosion (Bennetzen and Kellogg, 1997). Recent findings have shed some light on the mechanisms that are responsible for genome contraction. An excess of LTRs (the long terminal repeated sequences flanking retrotransposons) relative to the internal regions of the retrotransposon BARE-1 has been observed in barley. This suggested that recombination can occur between the LTRs, removing the internal domain and resulting in solo LTRs (Fig. 4; Vicient *et al.*, 1999; Shirasu *et al.*, 2000). Similarly, Wicker *et al.* (2001) found solo LTRs on a 211 kb wheat DNA sequence. They also observed more complex patterns of rearrangements probably involving a series of intra-element recombinations. In addition, evidence was found for the deletion of two large DNA fragments (up to 14 kb) consisting of different types of retroelements. Interestingly, this pattern of deletion could not be explained by intra-element recombination, suggesting that genome rearrangements counteracting genome expansion can occur independently of retrotransposon activity (Fig. 4; Wicker *et al.*, 2001). Other mechanisms that could also contribute to differences in genome size, such as a difference in the rate of DNA loss, have not yet been studied in plants. In insects, Petrov *et al.* (2000) have shown that in the Hawaiian cricket with a genome 11 times larger than that of *Drosophila melanogaster*, DNA is lost 40 times slower than in *D. melanogaster*.

Comparative genomic analysis in grasses increases understanding of the structure and evolution of grass genomes. With the development of new genomic tools, large stretches of DNA can now be sequenced in different grass species and compared. Comparison of such sequences with those of the rice model genome will aid detection of new mechanisms involved in grass genome evolution. The number and the diversity of the grass genomes that are currently under investigation represent a major advantage in comparative genomics. Map-based cloning approaches for the isolation of agronomically important genes as well as fundamental research (e.g. on polyploidization and on the dynamics of repetitive elements in shaping the genomes) will benefit from these studies.

ACKNOWLEDGEMENTS

We thank Dr C. Ringli and Dr N. Yahiaoui for critical reading of the manuscript.

LITERATURE CITED

- Asnaghi C, Paulet F, Kaye C, Grivet L, Deu M, Glaszmann JC, D'Hont A. 2000. Application of synteny across Poaceae to determine the map location of a sugarcane rust resistance gene. *Theoretical and Applied Genetics* **101**: 962–969.
- Avramova Z, Tikhonov A, SanMiguel P, Jin YK, Liu C, Woo SS, Wing R, Bennetzen JL. 1996. Gene identification in a complex chromosomal continuum by local genomic cross-referencing. *Plant Journal* **10**: 1163–1168.
- Bennetzen JL. 1996. The contribution of retroelements to plant genome organization, function and evolution. *Trends in Microbiology* **4**: 347–353.
- Bennetzen JL. 2000a. Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. *Plant Cell* **12**: 1021–1029.
- Bennetzen JL. 2000b. Transposable elements contributions to plant gene and genome evolution. *Plant Molecular Biology* **42**: 251–269.
- Bennetzen JL, Kellogg EA. 1997. Do plants have one way ticket to genomic obesity? *Plant Cell* **9**: 1507–1514.
- Bennetzen JL, SanMiguel P, Chen M, Tikhonov A, Francki M, Avramova Z. 1998. Grass genomes. *Proceedings of the National Academy of Sciences of the USA* **95**: 1975–1978.
- Bonierbale MD, Plaisted RL, Tanksley SD. 1988. RFLP maps on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* **120**: 1095–1103.
- Chen M, SanMiguel P, Bennetzen JL. 1998. Sequence organization and conservation in *sh2/a1*-homologous regions of sorghum and rice. *Genetics* **148**: 435–443.
- Chen M, SanMiguel P, De Oliveira AC, Woo SS, Zhang H, Wing RA, Bennetzen JL. 1997. Microcolinearity in *sh2*-homologous regions of the maize, rice and sorghum genomes. *Proceedings of the National Academy of Sciences of the USA* **94**: 3431–3435.
- Devos KM, Gale MD. 1997. Comparative genetics in the grasses. *Plant Molecular Biology* **35**: 3–15.
- Devos KM, Gale MD. 2000. Genome relationships: the grass model in current research. *Plant Cell* **12**: 637–646.
- Devos KM, Beales J, Nagamura Y, Sasaki T. 1999. Arabidopsis-rice: will colinearity allow gene prediction across the eudicot-monocot divide? *Genome Research* **9**: 825–829.
- Druka A, Kudrna D, Han F, Kilian A, Steffenson B, Frisch, Tomkins J, Wing R, Kleinohfs A. 2000. Physical mapping of the barley stem rust resistance gene *rpg4*. *Molecular Genetics and Genomics* **264**: 283–290.
- Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan LL, Shiloff BA, Bennetzen JL. 2001. Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. *Plant Physiology* **125**: 1342–1353.
- Feuillet C, Keller B. 1999. High gene density is conserved at syntenic loci of small and large grass genomes. *Proceedings of the National Academy of Sciences of the USA* **96**: 8665–8670.
- Feuillet C, Penger A, Gellner K, Mast A, Keller B. 2001. Molecular evolution of receptor-like kinase genes in hexaploid wheat: independent evolution of orthologs after polyploidization and mechanisms of local rearrangements at paralogous loci. *Plant Physiology* **125**: 1304–1313.
- Gale MD, Devos KM. 1998. Comparative genetics in the grasses. *Proceedings of the National Academy of Sciences of the USA* **95**: 1971–1974.
- Gaut BS, Le Thierry D'Ennequin M, Peek AS, Sawkins MC. 2000. Maize as a model for the evolution of plant nuclear genomes. *Proceedings of the National Academy of Sciences of the USA* **97**: 7008–7015.
- Keller B, Feuillet C. 2000. Colinearity and gene density in grass genomes. *Trends in Plant Science* **5**: 246–251.
- Kellogg EA. 1998. Relationships of cereal crops and other grasses. *Proceedings of the National Academy of Sciences of the USA* **95**: 2005–2010.
- Kellogg EA. 2001. Evolutionary history of the grasses. *Plant Physiology* **125**: 1198–1205.
- Kilian A, Chen J, Han F, Steffenson B, Kleinohfs A. 1997. Towards map-based cloning of the barley stem rust resistance gene *Rpg1* and *rpg4* using rice as a intergenomic cloning vehicle. *Plant Molecular Biology* **35**: 187–195.
- Kumar A, Bennetzen JL. 1999. Plant retrotransposons. *Annual Review of Genetics* **33**: 479–532.
- Langdon T, Seago C, Mende M, Leggett M, Thomas H, Forster JW, Thomas H, Neil Jones R, Jenkins G. 2000. Retrotransposons evolution in diverse plant genomes. *Genetics* **156**: 313–325.
- Lijavetzky D, Muzzi G, Wicker T, Keller B, Wing R, Dubcovsky J. 1999. Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. *Genome* **42**: 1176–1182.
- Moore G, Devos KM, Wang Z, Gale MD. 1995. Grasses, line up and form a circle. *Current Biology* **5**: 737–739.
- Moulet O, Zhang HB, Lagudah ES. 1999. Construction and characterisation of a large DNA insert library from the D genome of wheat. *Theoretical and Applied Genetics* **99**: 305–313.
- Panstruga R, Büschges R, Piffanelli P, Schulze-Lefert P. 1998. A

- contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. *Nucleic Acids Research* **26**: 1056–1062.
- Paterson AH, Lin YR, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE.** 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* **269**: 1714–1718.
- Paterson AH, Lan TH, Reischmann KP, Chang C, Lin YR, Liu SC, Burow MD, Kowalski SP, Katsar CS, Delmonte TA, et al.** 1996. Towards a unified genetic map of higher plants, transcending the monocot-dicot divergence. *Nature Genetics* **14**: 380–382.
- Peng AH, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, et al.** 1999. ‘Green revolution’ genes encode mutant gibberellin response modulators. *Nature* **400**: 256–261.
- Pereira MG, Lee M.** 1995. Identification of genomic regions affecting plant height in sorghum and maize. *Theoretical and Applied Genetics* **90**: 380–388.
- Petrov DA, Sangster TA, Spencer Johnston J, Hartl DL, Shaw KL.** 2000. Evidence for DNA loss as a determinant of genome size. *Science* **287**: 1060–1062.
- Roberts MA, Reader SM, Dalgliesh C, Miller TE, Foote TN, Fish LJ, Snape JW, Moore G.** 1999. Induction and characterization of *Phl* wheat mutants. *Genetics* **153**: 1909–1918.
- Sandhu D, Champoux JA, Bondavera SN, Gill KS.** 2001. Identification and physical localization of useful genes and markers to a major gene-rich region on wheat group 1S chromosomes. *Genetics* **157**: 1735–1747.
- SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL.** 1998. The paleontology of intergene retrotransposons of maize: dating the strata. *Nature Genetics* **20**: 43–45.
- SanMiguel P, Tikhonov A, Young-Kwan J, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL.** 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**: 765–768.
- Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P.** 2000. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Research* **10**: 908–915.
- Stein N, Feuillet C, Wicker T, Schlagenhauf E, 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. *Proceedings of the National Academy of Sciences of the USA* **97**: 13436–13441.
- Tarchini R, Biddle P, Wineland R, Tingey S, 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.
- Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JF, Avramova Z.** 1999. Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. *Proceedings of the National Academy of Sciences of the USA* **96**: 7409–7414.
- Van Deynze AE, Sorrells ME, Park WD, Ayres NM, Fu H, Cartinhour SW, Paul E, McCouch SR.** 1998. Anchor probes for comparative mapping of grass genera. *Theoretical and Applied Genetics* **97**: 356–369.
- van Dodeweerd AM, Hall C, Bent EG, Johnson SJ, Bevan MW, Bancroft I.** 1999. Identification and analysis of homoeologous segments of the genomes of rice and *Arabidopsis thaliana*. *Genome* **42**: 887–892.
- Vicient CM, Jääskeläinen MJ, Kalendar R, Schulman AH.** 2001. Active retrotransposons are common feature of grass genomes. *Plant Physiology* **125**: 1283–1292.
- Vicient CM, Suoniemi A, Anamthawat-Jonsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH.** 1999. Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum*. *Plant Cell* **11**: 1769–1784.
- Vision TJ, Brown DG, Tanksley SD.** 2000. The origins of genomic duplications in *Arabidopsis*. *Science* **290**: 2114–2117.
- Wang ZM, Le Thierry d’Ennequin M, Panaud M, Gale MD, Sarr A, Devos KM.** 2001. Trait mapping in foxtail millet. *Theoretical and Applied Genetics* (in press).
- Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B.** 2001. Analysis of a contiguous 211 kb sequence in diploid wheat (*T. monococcum* L.) reveals multiple mechanisms of genome evolution. *Plant Journal* **26**: 307–316.