

GENOMICS ARTICLE

Genome Relationships: The Grass Model in Current Research

Katrien M. Devos¹ and Mike D. Gale

John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, United Kingdom

INTRODUCTION

Ten years ago, with the advent of comparative mapping, a new tool became available to plant geneticists. Comparative genome analyses demonstrated that gene orders among related plant species remained largely conserved over millions of years of evolution. This finding has revolutionized our thinking and formed the basis of a new science—comparative genomics. Since the first comparative mapping experiments, studies of genome relationships have included an ever wider range of plant species, and the focus has shifted from comparisons at the gross map level to studies of gene organization in small chromosomal regions and finally to the DNA sequence itself.

To date, the most comprehensive data set comes from the grasses. The Poaceae family includes the staple cereals rice, maize, wheat, barley, sorghum, and the millets. For most of the past century, research has proceeded on each of these species individually. Mutants were produced, genetic maps were generated, and biochemical pathways were unraveled. Comparative genomics has provided the basis and stimulus to integrate this knowledge for application to all cereal crops.

The realization of the power of a comparative approach has led to several plant genome initiatives and has promoted rice as a model for cereal crops. The rice genome, with only 400 million DNA base pairs (Mb), is the smallest among the major cereal crops and only approximately four times larger than that of the eudicot model plant *Arabidopsis*. Dense genetic maps, carrying some 2500 markers, have been constructed, and physical maps cover most of the genome. Additionally, large public collections of expressed sequence tags (ESTs) have become available, and it is clear that sequence data covering the entire rice genome will add a further vital dimension to comparative genomics in the near future. Views on the prospective applications of comparative genomics have developed as new insights have been gained into genome relationships and genome organization. At the same time, new opportunities have been en-

visaged and exploited. Below, we present an overview of the status of comparative genomics, with examples of ongoing research projects, focusing on wheat.

GENOME COMPARISONS: FROM MAPS TO SEQUENCES

Beginning with a Map

Cereal crop species vary greatly in their DNA content, from ~400 Mb in the small rice and foxtail millet genomes to 17,000 Mb in bread wheat. The bulk of the larger genomes consists of repetitive DNA sequences, most of which are species specific. Genes, on the other hand, tend to be highly conserved at the DNA sequence level. This conservation allows the use of heterologous probes in DNA gel blot experiments to identify orthologous DNA sequences even in species belonging to different tribes within the same taxonomic family. Therefore, cross-mapping of gene sequences is the first step to study genome relationships.

Hexaploid Wheat

An evolutionary analysis of closely related genomes is one immediate result of molecular mapping within allopolyploid species. Wheat, for example, arose through spontaneous hybridization of the diploid species *Triticum urartu* (AA genome) and an unknown species (BB genome), probably belonging to the Sitopsis group of *Aegilops* spp, to form tetraploid wheat (AABB). Further hybridization with a third ancestral species, *Aegilops tauschii* (DD), led, ~10,000 years ago, to the production of hexaploid bread wheat ($2n = 6x = 42$; AABBDD). The finding that the three wheat genomes have a highly similar gene content and order was the first demonstration of colinearity in the grasses and a pivotal finding in the development of comparative genomics (Chao et al., 1989). Subsequently, it has become clear that the

¹To whom correspondence should be addressed. E-mail katrien.devos@bbsrc.ac.uk; fax 44-1603-450024.

genomes are rearranged, relative to one another, through the occurrence of large reciprocal translocations involving chromosome arms 2BS and 6BS (Devos et al., 1993b) and chromosomes 4A, 5A, and 7B (Devos et al., 1995). The first translocation in this latter complex involved chromosome arms 4AL and 5AL and took place in the diploid ancestor. The remaining rearrangements probably arose at the tetraploid level.

The Triticeae Tribe

Comparisons at the genetic map level within the tribe Triticeae, which includes wheat, barley, rye, and wild relatives such as *Aegilops* spp, again show a high conservation of genome colinearity, which is only disrupted by gross chromosomal translocations. Interestingly, the number of rearrangements observed between different genomes varies greatly and appears to be unrelated to phylogenetic distance or evolutionary time. For example, the barley (*Hordeum vulgare*) genome is highly colinear with that of *A. tauschii*, differing only by two possible inversions (Dubcovsky et al., 1996), whereas the *A. umbellulata* and *A. tauschii* genomes differ by a minimum of 11 rearrangements (Zhang et al., 1998; Figure 1). Phylogenetic data have, however, indicated that *A. umbellulata* is more closely related to *A. tauschii* than is barley (Kellogg et al., 1996). Analyses of other Triticeae, such as rye (*Secale cereale*) (Devos et al., 1993a), *A. longissima* (Naranjo, 1995; Zhang et al., 2000), and *A. speltoides* (Maestra and Naranjo, 1998), provide further evidence that some genomes fix rearrangements more readily than others (Figure 1). These different rates of species divergence through chromosomal rearrangement do not appear to correlate to the breeding system, because high levels of evolutionary translocations are found in both rye, an outbreeder, and *A. umbellulata*, a predominantly self-pollinated species.

A comparison of centromere positions allows us to infer the basic structure of the "ancestral" Triticeae chromosomes. Kimber (1967) observed the centromeres in *Triticum* and *Aegilops* spp to be generally located in submedian positions and suggested that subterminal centromere positions were the result of chromosomal rearrangements. The comparative maps indeed show that the subterminal positions of *A. umbellulata* chromosomes 2U and 7U are due to inversions and those of 3U and 6U to interchromosomal translocations (Figure 1).

The Poaceae Family

Gene content and order are also well maintained within taxonomic families, despite large differences between species in genome size and chromosome number. Cross-mapping of restriction fragment length polymorphism probes in a range of cereal crops has led to the widespread identification of chromosomal regions in which marker orders are

highly conserved. It is now possible to describe all the genomes of grass crop species by their relationship to a single reference genome, rice. Regions of the rice genome that contain sets of markers that are colinear across other grass species have been termed linkage blocks.

Figure 2 depicts the genomes of foxtail millet (*Setaria italica*; $2n = 2x = 18$; $C = 0.45$ pg) and pearl millet (*Pennisetum glaucum*; $2n = 2x = 14$; $C = 2.4$ pg), two species belonging to the Paniceae tribe within the Panicoideae subfamily, in relation to rice (*Oryza sativa*; $2n = 2x = 24$; $C = 0.4$ pg). The highly similar arrangement of linkage blocks in foxtail millet and rice demonstrates that these genomes have undergone few rearrangements since their divergence. The pearl millet genome, on the other hand, is highly rearranged relative to rice. Many of these rearrangements are not present in any of the other Panicoideae analyzed to date. Therefore, they are likely to be specific either to pearl millet or to the genus *Pennisetum*. A few chromosomal rearrangements, however, can be identified that are likely to have occurred before the speciation of the Panicoideae but after the Panicoideae–Oryzoideae divergence. Examples are the organization of rice linkage block 10, which is inserted into rice linkage block 3, and the insertion of rice linkage block 9 into 7 to form specific Panicoideae chromosomes (Figure 2). A different arrangement of rice linkage blocks, the insertion of linkage block 8 into 6 and linkage block 10 into 5, is present in wheat and oat and may characterize the Pooideae subfamily (Gale and Devos, 1998).

Some rearrangements may reflect more ancient evolutionary events. A particularly interesting observation is that regions duplicated on rice chromosome arms 11S and 12S (Nagamura et al., 1995) have orthologs on foxtail millet chromosomes VII and VIII and pearl millet linkage groups 1 and 4 (Figure 2). The duplication event must therefore predate the divergence of the Panicoideae and rice from a common ancestor. This duplication has not been observed in members of the Pooideae subfamily; however, this does not preclude its presence even there. Duplications may go unnoticed when the marker density in the critical chromosome regions is low or when low polymorphism levels prohibit the mapping of both copies of duplicated probes.

To date, genome relationships have been established between rice, wild rice (*Zizania palustris*; Kennard et al., 1999), foxtail millet, sugar cane, sorghum, pearl millet, maize, the Triticeae cereals wheat, barley, rye, and oats (for overview, see Gale and Devos, 1998), and a growing number of wild relatives of the domesticated cereal crop species. Despite a divergence time of some 60 million years, less than 30 rice linkage blocks are needed to represent all of these genomes. Significantly, the boundaries of the rice linkage blocks frequently coincide with the location of centromeres and telomeres, which implies that these sites may play a key role in chromosome evolution (Qumsiyeh, 1994; Moore et al., 1997). Other cereals that will be integrated into the grass consensus map in the near future include finger millet (*Eleusine coracana*) and tef (*Eragrostis tef*), two members of

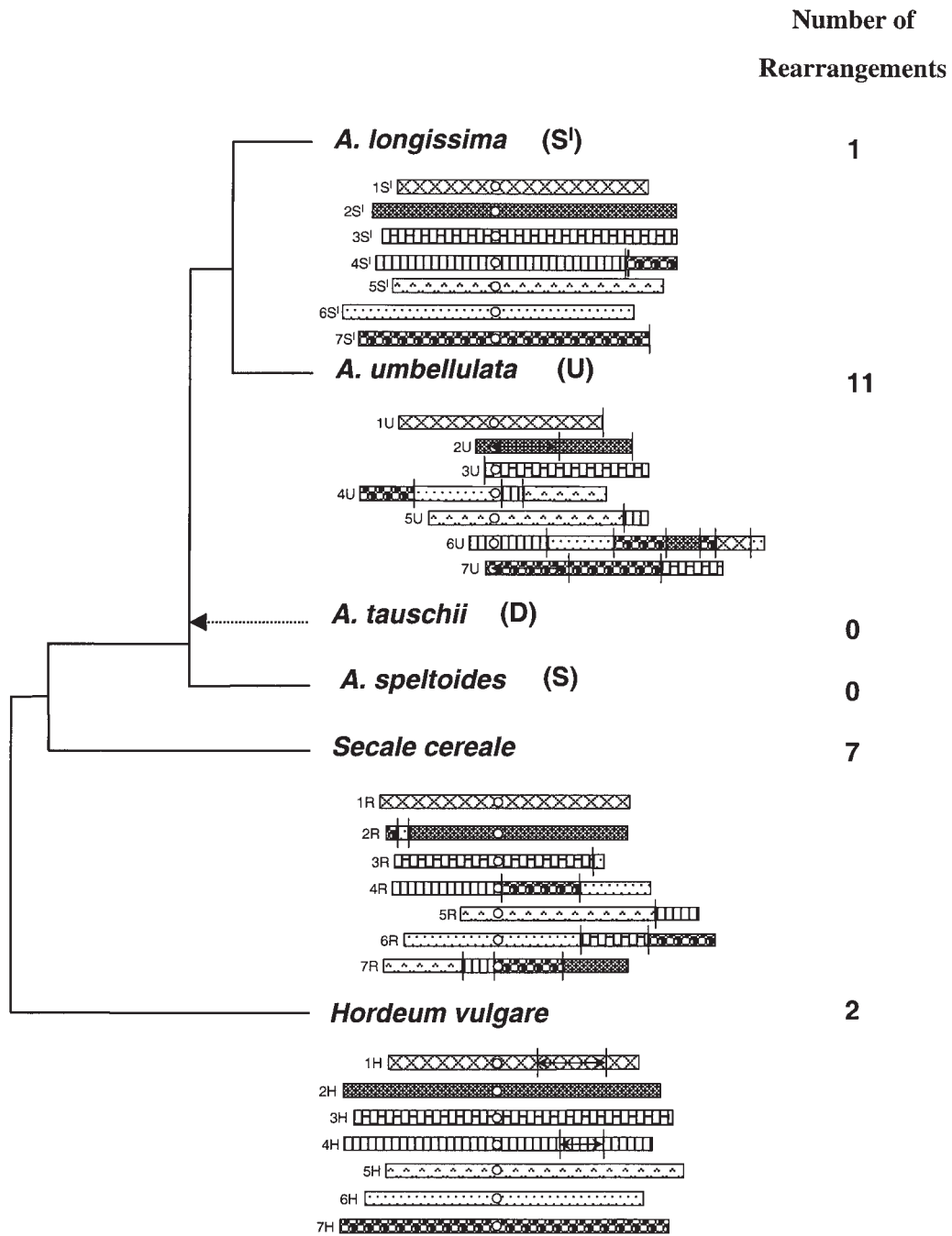


Figure 1. Phylogenetic and Genome Relationships among Six Triticeae Grasses.

The D genome of *A. tauschii* has been chosen as the reference, and each of the seven chromosomes is represented by a different pattern. Open circles indicate centromere positions. Double-headed arrows indicate inversions. Vertical lines indicate evolutionary translocation breakpoints. The minimum number of chromosomal rearrangements necessary for the six present-day genomes to have evolved from the D genome format is given for each species.

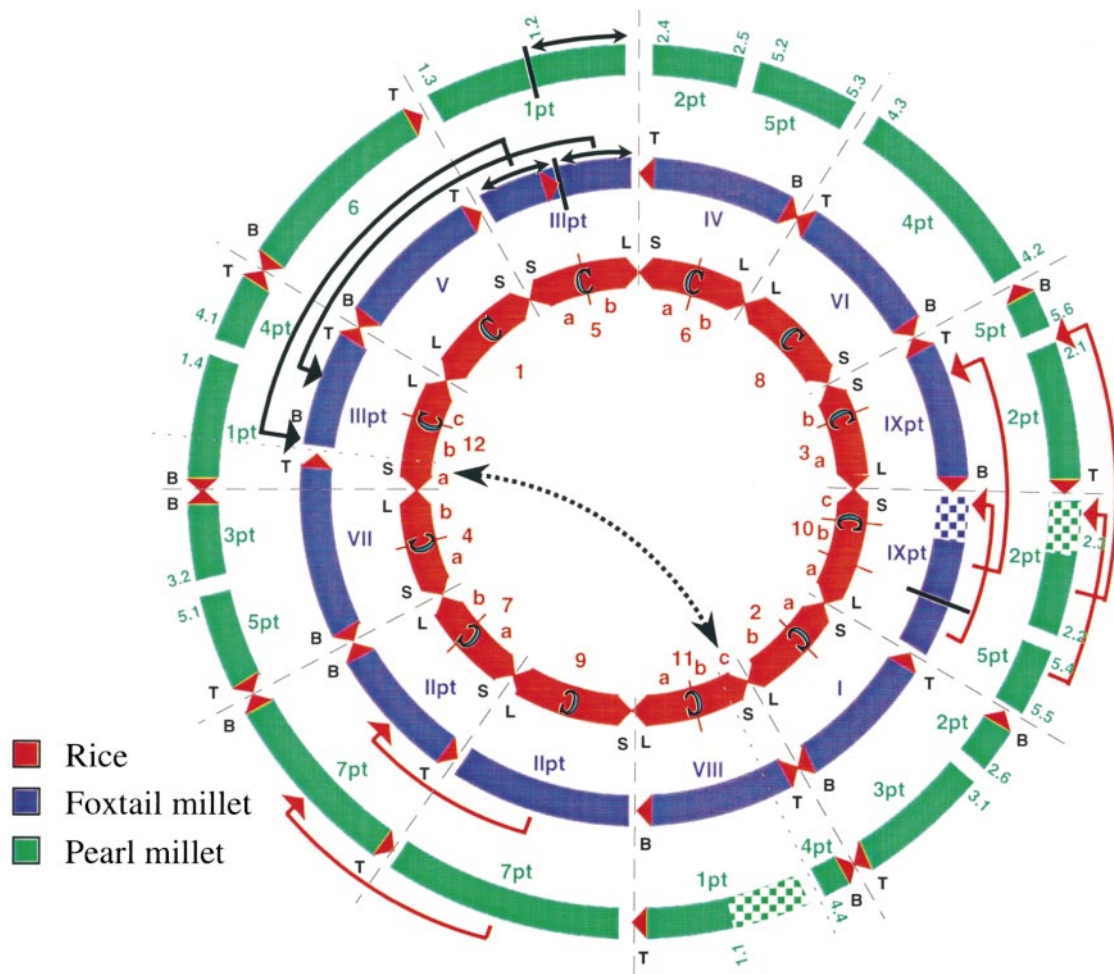


Figure 2. Relationships among the Genomes of Rice, Foxtail Millet, and Pearl Millet.

C indicates rice centromere positions. Red triangles indicate telomeres. S and L indicate short and long arms, respectively, as assigned to the rice chromosomes, which are numbered with arabic numerals. Rice linkage blocks correspond to entire chromosomes (e.g., rice chromosomes 1, 8, and 9) or chromosome segments (e.g., 2a, 2b, 3a, and 3b). In foxtail and pearl millet, for which linkage groups have not been assigned to short and long arms, the designations top (T) and bottom (B), which correspond to the published maps (Devos et al., 1998, 2000), are used. Foxtail millet chromosomes are numbered with roman numerals, and pearl millet linkage groups are numbered with arabic numerals. For chromosomes that have orthology to more than one rice chromosome, segments are indicated with the chromosome number followed by pt (part). Hatched areas indicate regions for which few comparative data are available. Double-headed arrows show inversions; single-headed arrows denote evolutionary translocations. In pearl millet, due to the large number of rearrangements relative to rice, the majority of arrows are omitted. Rearrangements can be derived from the chromosome segment numbers; for example, pearl millet linkage group 5 is orthologous from top to bottom (T-5.1/5.2-5.3/5.4-5.5/5.6-B) with rice linkage blocks 4a/6b/10a/3b. Red arrows show evolutionary translocations that characterize all *Panicoidae* spp analyzed to date. The dotted arrow indicates the rice 11S/12S duplication. See the text for details.

the Chloridoideae subfamily, and *Lolium* and *Festuca* spp, two members of the Pooidae subfamily. The Chloridoideae are a new subfamily not previously represented. In addition to assessing the overall levels of colinearity, it will be of interest to identify chromosomal arrangements that are present in both tef and finger millet. The presence of common rearrangements between subfamilies may provide evidence for taxonomic relationships. Preliminary data on the

structural organization of the tetraploid finger millet genome have indicated the presence of the ancient interchromosomal R11S/R12S duplication previously observed in rice, foxtail, and pearl millet (M.M. Dida and K.M. Devos, unpublished data). Therefore, this duplication is predicted to be present also in tef.

The comparative framework of molecular markers can be used for map-based prediction of the location of genes that

determine key traits. Excellent examples are those genes controlling plant height that are insensitive to gibberellin application. In wheat, the *Rht-1* genes are located on the short arms of chromosomes 4B and 4D, which correspond to the duplicated regions on maize chromosome arms 1L and 5S carrying the gibberellin-insensitive dwarfing genes *D8* and *D9*. Isolation of *Rht-1* and *D8* has confirmed that the maize and wheat genes are indeed orthologous. In both cases, dwarf phenotypes were caused by mutations that altered the N-terminal region of the encoded proteins (Peng et al., 1999).

Although there are few cases in which the orthology of genes underlying traits of agronomic importance has been confirmed by sequence information, the mapping of traits to orthologous regions provides a reasonable degree of certainty as to the evolutionary descent of such genes. The domestication of crops, for instance, has involved the modification of one or more genes affecting seed dispersal, and seed shattering in foxtail millet has recently been ascribed to two major quantitative trait loci (QTL) on chromosomes V and IX (Wang et al., 2000). Comparative analysis showed the QTL region on chromosome IX to correspond to regions of maize chromosomes 1 and 5 and sorghum linkage group C, which had previously been shown to carry genes controlling seed dispersal (Figure 3). Similarly, the QTL region on foxtail millet chromosome V was orthologous to regions of rice chromosome 1, pearl millet linkage group 6, and maize chromosomes 3 and 8, all of which carry genes controlling shattering ability (Paterson et al., 1995; Wang et al., 2000). Conserved colinearity is likely to hold up for most adaptive genes; however, disease resistance genes may be an exception in that they can undergo a more rapid reorganization (Leister et al., 1998).

Refining the Map: Megabase Resolution

Although colinearity at the map level can be used in taxonomy and as a predictive tool, comparative map-based gene isolation requires highly conserved gene orders at the 100-kb to 1-Mb level. Within the plant kingdom, monocot and eudicot species have similar numbers of genes. In maize, it has been demonstrated that genome expansion is mainly due to the relatively recent (within the last 3 to 6 million years) accumulation of retrotransposons (SanMiguel et al., 1996, 1998). In wheat and barley, ~80% of the genome consists of highly repetitive DNA, most of which is genome specific. Although these large amounts of repetitive DNA may make chromosome walking extremely difficult, potential problems may be circumvented by using a small-genome relative as a model, provided that the genes in the target region are present in almost precisely the same order as those in the larger reference genome. Sorghum, which has a haploid DNA content of 0.8 pg and is closely related to maize (1C = 2.8 pg), is an obvious candidate to aid genome analysis in maize. Detailed comparisons between maize and

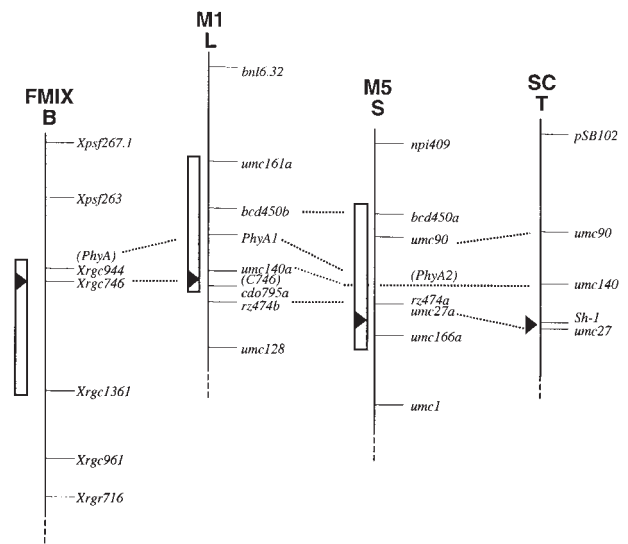


Figure 3. Orthologous Regions of Foxtail Millet, Maize, and Sorghum Chromosomes That Carry Genes Controlling Shattering of the Inflorescence.

The chromosome regions shown are the bottom of foxtail millet IX (FMIX, B), the long arm of maize 1 (M1, L), the short arm of maize 5 (M5, S), and the top of sorghum C (SC, T), which have been shown to carry genes controlling shattering. Rectangles delineate the extent of the QTL effect as observed or, in the case of sorghum, the position of a mapped single gene, *Sh-1*. Maize maps are composite maps based on the data by Paterson et al. (1995) and Davis et al. (1999).

sorghum at the megabase and sequence level are described by Bennetzen (2000).

Wheat (1C = 17 pg) has no small-genome relatives for which genetic data and tools are available. Therefore, the use of rice as a model has been investigated. Foote et al. (1997) constructed a detailed genetic map of a region of the wheat chromosome arm (5BL) that carries *Ph1*, the gene that controls homeologous chromosome pairing, and thereby maintains disomic inheritance in allohexaploid bread wheat. Linkage block analysis has revealed that the wheat *Ph1* region is orthologous to a region of rice chromosome 9. In a higher resolution study, the marker order in the *Ph1* region of wheat was compared with that in a rice yeast artificial chromosome (YAC) contig spanning the target region on rice chromosome 9. Although marker orders were generally conserved, a small segment spanning three markers was found to be duplicated in rice, with one copy in the target region and the second copy ~10 centimorgans distal from it. In addition, disruption of colinearity was observed outside the target region, suggesting that utility of rice for gene isolation in wheat may depend on the particular chromosome region. Currently, ~300 kb of rice genomic DNA

spanning the location of the rice *Ph1* homolog is being sequenced (Roberts et al., 1999).

Feuillet and Keller (1999) have investigated the barley, maize, and rice genomes for colinearity with a region of the wheat genome spanning the receptor-like kinase genes *Lrk10* and *Tak10*. This region was found to be duplicated in wheat, with one copy being located on the short arms of the group 1 chromosomes and the second copy on the short arms of the group 3 chromosomes. No orthologous regions were found in rice and maize for the wheat group 1 segment. A large family of genes related to *Lrk10* was identified distally on rice chromosome 1 and on maize chromosome 8, which corresponds to the *Lrk-Tak* region on the wheat group 3 chromosomes. Analysis of the number and organization of the *Lrk-Tak* genes in wheat, maize, and rice nevertheless showed that the orthologous segments differed by duplications and rearrangements (Feuillet and Keller, 1999).

A similar comparative analysis at the megabase level was conducted between barley and rice in a search for the barley stem rust resistance gene *Rpg1* (Kilian et al., 1997). A high-density map of the chromosome region around *Rpg1* was constructed, and gene order was generally found to be conserved relative to a rice contig spanning the target region. Again, as was observed in other studies, a few markers appeared to have transposed to nearby noncolinear positions. Moreover, a rice homolog of *Rpg1* could not be found within the bacterial artificial chromosome (BAC) containing the *Rpg1* flanking markers. Although the gene for a putative membrane protein could not be excluded as a candidate, it is likely that *Rpg1* is a rare exception to the conserved gene order of that region in barley and rice (Han et al., 1999). Thus, the emerging data suggest that the extent of colinearity observed between the Triticeae crops and rice may vary between chromosome regions and that even in regions in which gene orders are highly conserved, colinearity is not absolute. The gross chromosomal organization may have remained largely conserved for 60 million years, but small local rearrangements and duplications are clearly a common feature of genome evolution.

It should again be noted that some of the observed rearrangements in gene order may pertain in particular to disease resistance genes. This conclusion was also inferred by a comparative analysis of the *Adh1-Adh2* region of rice with maize. From 33 genes identified within a 350-kb stretch of rice genomic sequence, 13 were tested for their ability to cross-hybridize to maize. Four of these detected colinear orthologs on maize chromosome 4. *Adh1* itself mapped to a nonsyntenic position on maize chromosome 1, providing further evidence that minor structural changes have occurred in most chromosomal regions after divergence from a common ancestor. Eight rice genes, including several disease-resistance gene homologs, however, failed to hybridize to maize genomic DNA, indicating either substantial sequence divergence or absence of these genes in maize (Tarchini et al., 2000). Rapidly evolving gene families may thus not be amenable to comparative map-based gene

cloning. It is also possible that gene loss is better tolerated in maize because of the tetraploid nature of its genome. Tolerance of wholesale gene loss in aneuploids is well documented in hexaploid wheat (Sears, 1954).

From Rice to Arabidopsis

Whether the colinearity observed between genomes within the grass family—despite 60 million years of divergent evolution—extends to Arabidopsis, which diverged some 150 million years earlier from a common ancestor, is a particularly relevant question because a large proportion of the Arabidopsis genomic sequence is already available. The existence of colinearity between Arabidopsis and the grasses, even if limited to small regions, would allow direct exploitation of the Arabidopsis genomic sequence for the identification of candidate genes in the cereals. Paterson et al. (1996) suggested that 43 to 58% of those chromosomal tracts spanning <3 centimorgans should have remained colinear over the evolutionary time period separating the monocots and eudicots. Evidence from a recent study, however, has not supported this hypothesis. The mapping of 33 rice ESTs, identified through BLAST searches as putative homologs to Arabidopsis genes that were located on the same or closely linked BAC clones, failed to establish a region of colinearity between rice and Arabidopsis (Devos et al., 1999; Figure 4). A similar comparative analysis of a 1.5-Mb region from Arabidopsis chromosome 4, on the other hand, revealed limited orthology of a 194-kb region of the Arabidopsis genome with 219 to 300 kb of the rice genome (van Dodeweerd et al., 1999). However, only five out of the 24 rice ESTs that showed homology to Arabidopsis genes in the 252-kb segment analyzed mapped to a single 530-kb rice YAC clone. Moreover, the order of these five genes differed in rice from that in Arabidopsis by an inversion. The remaining 19 ESTs were located elsewhere in the rice genome and, because they did not hybridize to common BACs, were probably not tightly clustered (van Dodeweerd et al., 1999).

There are, of course, several problems associated with these analyses. BLAST algorithms, for example, are not impeccable in identifying true homologs, especially when given incomplete databases. The best BLAST correlation from a given search may represent an ortholog, or it may simply represent a member of a gene family. It may even identify a nonrelated gene that contains a similar sequence domain. Genes that contain a conserved MADS box domain, for example, may be otherwise unrelated (Theissen et al., 2000). To further assess the extent of colinearity between rice and Arabidopsis, therefore, we performed an analysis at the sequence level that included ~250 kb of contiguous sequence from rice chromosome 1 (<http://www.staff.or.jp/genomicdata/GenomeFinished.html>). For four out of the 51 rice genes located within this 250-kb region, the Arabidopsis gene with the highest BLAST score was located to a 350-kb Arabidopsis segment. An additional 10 rice

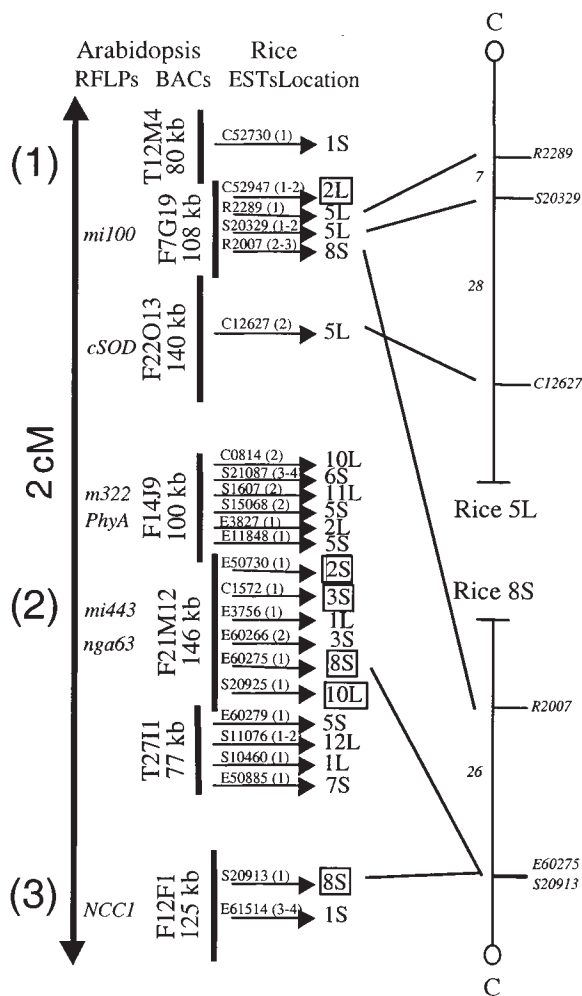


Figure 4. Syntenic Relationships between Arabidopsis Genes and Corresponding Rice ESTs.

A genetic map of an ~2-centimorgan (cM) region at the top of Arabidopsis chromosome 1 is aligned with seven Arabidopsis BACs organized in three contigs. Also shown are corresponding rice ESTs, their copy number, and map locations on rice chromosome arms 5L and 8S. For ESTs in which true homology between the rice EST and Arabidopsis gene has been established, the locations are shown in boxes. RFLP, restriction fragment length polymorphism.

genes displayed homology with Arabidopsis genes elsewhere in the genome. Interestingly, five of these 10 genes also showed homology with Arabidopsis genes within the 350-kb contig, albeit with a lower BLAST value (K.M. Devos, unpublished data). The ancestral relationship between these genes remains unclear at this stage, but it is doubtful that they are real orthologs.

Caution is needed in interpreting comparative data for genomes as divergent as those of rice and Arabidopsis. The

degree of colinearity found largely depends on the stringency of the parameters used. More important than the issue of colinearity, however, is the utility of the relationship between Arabidopsis and rice for map-based gene prediction. Based on current data, “conserved colinearity” applies at best to only a few genes within syntenic Arabidopsis–rice regions. Therefore, the prospect of scanning the Arabidopsis genome to find a gene for a trait of agronomic importance in monocot crop plants based on relative map position continues to be like searching for a needle in a haystack. Transferring markers that are tightly linked to and flank a key wheat gene to Arabidopsis is unlikely to delineate the Arabidopsis region carrying the candidate gene. Extrapolation between wheat and rice, on the other hand, entails a reasonable chance of success given the high level of conserved colinearity between their genomes. The sequencing of the entire rice genome is therefore crucial to investigations into the staple cereals in general.

Complications in Determining Synteny: Multigene Families and Chromosomal Duplications

A number of factors may confound interpretations of genome relationships. The use of programs such as BLAST, for example, is crucial in comparisons of genomes that are only distantly related. However, the identification of “homologous” genes based on BLAST results is dependent on the particular criteria used by the researcher. In addition, the differentiation between orthologous and paralogous sequences is extremely difficult, especially when both are not available for comparison.

Problems in distinguishing between orthologs, paralogs, and pseudogenes can also be encountered in comparisons of closely related species for which genome colinearity has been established through the cross-hybridization of conserved gene sequences. Only those probes that identify a single-copy gene for each of the genomes under investigation are likely to give unambiguous answers. Genes belonging to multigene families will often hybridize to many members of the gene family, but only those that display restriction fragment length polymorphism in the mapping population can be genetically mapped. When different members of a single family are mapped in different species, the impression is created that colinearity is disrupted. For example, all probes that detected noncolinear loci in a comparative study between the barley and *A. tauschii* genomes were multicopy in one or both of the species (Namuth et al., 1994). Similarly, in a comparison between foxtail millet and rice, of the probes that mapped to noncolinear positions, all but one detected at least two copies in the foxtail millet genome (Devos et al., 1998). It is possible to envisage a situation in which an apparently single-copy probe may detect nonorthologous sequences. After a duplication event in a common ancestor, one of the gene copies may subsequently be deleted in each of two species after their divergence from the common

ancestor. Deletion of paralogous copies leads to single-copy sequences that have noncolinear positions in the two genomes.

Segmental chromosome duplications also may confound relationships. In diploid rice, duplications have been identified between rice chromosome arms 11S and 12S (Nagamura et al., 1995) and segments of rice chromosomes 1 and 5 (Kishimoto et al., 1994). It is expected that more duplications will be identified as sequencing information becomes available. Indeed, the availability of the near-complete sequence of the Arabidopsis genome, for example, has revealed the presence of both intra- and interchromosomal duplications. Although considered a true diploid, >60% of the predicted proteins on Arabidopsis chromosome 2 were shown to have a significant match to at least one other protein on the same chromosome. In addition, large interchromosomal duplications have been identified between two segments of chromosomes 1 and 2 and between regions of chromosomes 2 and 4 (Lin et al., 1999). Gene multiplication is a common event in genome evolution, although it is not clear at this stage whether this occurs mainly through duplication of individual genes or larger chromosome segments. After duplication, redundant gene copies are likely to be more prone to accumulate mutations, which may eventually lead to altered gene functions. It is probable that all living organisms will have undergone a certain level of chromosomal duplications.

ORGANIZATION OF THE WHEAT AND BARLEY GENOMIC SEQUENCES

Species with large genomes, such as the Triticeae cereals, have long been deemed unamenable to gene isolation by map-based methods. More than 80% of the wheat and barley genomes consist of highly repetitive DNA, which may cause problems in the construction as well as analysis of large insert libraries. YACs containing repeats are often unstable, and isolation of single-copy YAC end or subclones suitable for chromosome walking is difficult (Dunford et al., 1993; Edwards et al., 1996). With the advent of comparative genomics and the realization that gene orders are highly conserved among grass genomes, the small rice genome has become a tool for cross-genome gene isolation. It is clear, however, that although levels of colinearity are in general highly conserved, the degree of gene order conservation may vary, depending on the region, and is unlikely to be perfect. So, is comparative genomics the only option?

Recent advances in the construction of large insert libraries, particularly in the development of BAC vectors, which have largely replaced YACs, and the availability of enhanced robotics capability have made the construction and maintenance of stable libraries of the entire wheat genome feasible. Currently, BAC libraries are available for several Triticeae species, including *T. monococcum* (Lijavetzky et

al., 1999), *A. tauschii* (Moulet et al., 1999), and barley (http://www.genome.clemson.edu/lib_frame.html). A barley YAC library is also available, and a hexaploid wheat library has been constructed in a transformation-competent artificial chromosome (TAC) vector (Ogihara et al., 2000; <http://www.intl-pag.org/pag/8/abstracts/pag81031.html>).

Physical mapping of molecular markers using a series of wheat deletion lines had demonstrated the existence of gene-rich regions in the wheat genome, indicating that genes were not randomly distributed (Gill et al., 1996). A more precise picture of the detailed organization of the wheat genome has been provided by analysis of large insert clones. Hybridization of an *A. tauschii* BAC clone containing ~100 to 105 kb around the *Cre3* locus (i.e., from the distal region of chromosome 2D) to a cDNA library identified six genes (O. Moulet and E.S. Lagudah, personal communication), thereby establishing a gene density around the *Cre3* locus of at least one gene every 17 kb, and possibly much less. In other regions, gene densities varied from one gene per 33 kb for the proximal region of chromosome 4DL (W. Powell, O. Moulet, and E.S. Lagudah, personal communication) and one gene per 25 kb around the *Ha* locus on 5AS (Tranquilli et al., 1999) to as high as one gene per 5 kb around the *Lrk10* locus on 1AS (Feuillet and Keller, 1999). The observed gene density in barley was very similar, with a value of one gene per 20 kb around the *Mlo* gene (Panstruga et al., 1998) and one gene per 15 kb around the *Lrk-Tak* loci (Feuillet and Keller, 1999). These values are seven to 40 times higher than expected based on a random gene distribution, and they agree with the hypothesis that genes are organized in islands.

Analysis of 60 kb of contiguous sequence around the *Mlo* gene in barley showed the presence of three genes, which made up 11.7% of the region. Interestingly, only a single retrotransposon was found, accounting for 24% of the region (Panstruga et al., 1998). This is very different from the structural organization of the distinctive maize genome, in which genes are often located between retroelements or retroelement blocks. Another detailed analysis of 340 kb of rice genomic sequence revealed that 28.5% of the region consisted of repetitive DNA, half of which was represented by retrotransposons and ~18.8% by miniature inverted repeat transposable elements. In contrast to the situation in maize, no clustering of the retroelements was evident (Tarchini et al., 2000). As the genomic sequence begins to emerge, gene densities in rice are being found to be of the order of one gene every 5 to 10 kb (<http://www.staff.or.jp/genomicdata/GenomeFinished.html>; Han et al., 1999; Tarchini et al., 2000), which is also 1.5 to 2.5 times higher than expected.

Sequence data from plant species with large genomes are still very limited. All available data suggest that genes in wheat and barley are clustered within the genome. If these gene islands can be targeted, then direct gene isolation should be greatly facilitated. The current approach for gene isolation is to identify molecular markers that are tightly

linked to a target trait in large populations (1000 to 10,000 plants). Markers tightly linked to and flanking the gene are then used to identify BAC or YAC clones spanning the gene. With gene densities of approximately one gene per 20 kb, it should be feasible to have both flanking markers located on a BAC contig of a few hundred kilobases, if not within a single BAC clone (Büsches et al., 1997; Wei et al., 1999).

MODEL PLANTS VERSUS CROP PLANTS

Over the years, there has been a shift in our approach to analyzing large genomes. Fifteen years ago, large genomes were considered essentially intractable. This view has changed during the 1990s with the development of comparative genetics. The high degree of conservation of gene content and order between species within taxonomic groups allowed the use of species with smaller genomes as models for crop traits. Now, at the turn of the century, scientists are faced with another dilemma. As more data become available, it is clear that the efficiency with which model organisms can be employed will depend on our knowledge of the trait itself and the genome region carrying the gene. It is thus pertinent to ask whether the exploitation of conserved colinear relationships is the best approach to crop plant research or whether we should focus research on the crop plant genomes themselves. The recent advances in technology and our knowledge of the structural organization of crop genomes definitely make the latter option feasible, as exemplified by the isolation of the barley *Mlo* gene. On the other hand, the isolation of the wheat *Rht1* dwarfing genes, using the Arabidopsis gene *GAI*, underlines the contribution that model species can make to crop plant research. Whatever route is taken, comparative genomics will continue to play a major role. The sequencing of the entire rice genome and selected regions of other crop genomes will provide a ready pool of candidate genes relevant to both agronomic and evolutionary considerations of grass genomes. It is important, however, that the progress in genome research be paralleled by developments in trait mapping and bioinformatics. Maximum exploitation of the vast pool of genome data will only be possible when suitable management systems and search and display tools are available.

REFERENCES

- Bennetzen, J.L. (2000). Comparative sequence analysis of plant nuclear genomes: Microcolinearity and its many exceptions. *Plant Cell* **12**, in press.
- Büsches, R., et al. (1997). The barley *Mlo* gene: A novel control element of plant pathogen resistance. *Cell* **88**, 695–705.
- 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.
- Davis, G.L., et al. (1999). A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. *Genetics* **152**, 1137–1172.
- Devos, K.M., Atkinson, M.D., Chinoy, C.N., Harcourt, R.L., Koebner, R.M.D., Liu, C.J., Masojc, P., Xie, D.X., and Gale, M.D. (1993a). Chromosomal rearrangements in the rye genome relative to that of wheat. *Theor. Appl. Genet.* **85**, 673–680.
- Devos, K.M., Millan, T., and Gale, M.D. (1993b). Comparative RFLP maps of the homoeologous group-2 chromosomes of wheat, rye and barley. *Theor. Appl. Genet.* **85**, 784–792.
- Devos, K.M., Dubcovsky, J., Dvorák, J., Chinoy, C.N., and Gale, M.D. (1995). Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. *Theor. Appl. Genet.* **91**, 282–288.
- Devos, K.M., Wang, Z.M., Beales, J., Sasaki, T., and Gale, M.D. (1998). Comparative genetic maps of foxtail millet (*Setaria italica*) and rice (*Oryza sativa*). *Theor. Appl. Genet.* **96**, 63–68.
- 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.
- Dubcovsky, J., Luo, M.-C., Zhong, G.-Y., Bransteitter, R., Desai, A., Kilian, A., Kleinhofs, A., and Dvorák, J. (1996). Genetic map of diploid wheat, *Triticum monococcum* L. and its comparison with maps of *Hordeum vulgare* L. *Genetics* **143**, 983–999.
- Dunford, R., Vilageliu, L., and Moore, G. (1993). Stabilization of a yeast artificial chromosome containing plant DNA using a recombination-deficient host. *Plant Mol. Biol.* **21**, 1187–1189.
- Edwards, K.J., Veuskens, J., Rawles, H., Daly, A., and Bennetzen, J.L. (1996). Characterization of four dispersed repetitive DNA sequences from *Zea mays* and their use in constructing contiguous DNA fragments using YAC clones. *Genome* **39**, 811–817.
- 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.
- Foote, T., Roberts, M., Kurata, N., Sasaki, T., and Moore, G. (1997). Detailed comparative mapping of cereal chromosome regions corresponding to the Ph1 locus in wheat. *Genetics* **147**, 801–807.
- Gale, M.D., and Devos, K.M. (1998). Plant comparative genetics after 10 years. *Science* **282**, 656–659.
- Gill, K.S., Gill, B.S., Endo, T.R., and Boyko, E.V. (1996). Identification and high density mapping of gene rich regions in chromosome group 5 of wheat. *Genetics* **143**, 1001–1012.
- 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.
- Kellogg, E.A., Appels, R., and Mason-Gamer, R.J. (1996). When genes tell different stories: The diploid genera of Triticeae (Gramineae). *Syst. Bot.* **21**, 1–17.

- Kennard, W., Phillips, R., Porter, R., and Grombacher, A. (1999). A comparative map of wild rice (*Zizania palustris* L. $2n=2x=30$). *Theor. Appl. Genet.* **99**, 793–799.
- 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.
- Kimber, G. (1967). The addition of the chromosomes of *Aegilops umbellulata* to *Triticum aestivum* (var. "Chinese Spring"). *Genet. Res.* **9**, 111–114.
- Kishimoto, N., Higo, H., Abe, K., Arai, S., Saito, A., and Higo, K. (1994). Identification of the duplicated segments in rice chromosomes 1 and 5 by linkage analysis of cDNA markers of known functions. *Theor. Appl. Genet.* **88**, 722–726.
- Leister, D.M., Kurth, J., Laurie, D.A., Yano, M., Sasaki, T., Devos, K.M., Graner, A., and Schulze-Lefert, P. (1998). Rapid reorganization of resistance gene homologues in cereal genomes. *Proc. Natl. Acad. Sci. USA* **95**, 370–375.
- Lijavetzky, D., Muzzi, G., Wicker, T., Keller, B., Wing, R., and Dubcovsky, J. (1999). Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. *Genome* **42**, 1176–1182.
- Lin, X.Y., et al. (1999). Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **402**, 761–768.
- Maestra, B., and Naranjo, T. (1998). Homoeologous relationships of *Aegilops speltoides* chromosomes to bread wheat. *Theor. Appl. Genet.* **97**, 181–186.
- Moore, G., Roberts, M., Aragon-Alcaide, L., and Foote, T. (1997). Centromeric sites and cereal chromosome evolution. *Chromosoma* **105**, 321–323.
- Moulet, O., Zhang, H.B., and Lagudah, E.S. (1999). Construction and characterisation of a large DNA insert library from the D genome of wheat. *Theor. Appl. Genet.* **99**, 305–313.
- Nagamura, Y., Inoue, T., Antonio, B.A., Shimano, T., Kajiya, H., Shomura, A., Lin, S.Y., Kuboki, Y., Harushima, Y., Kurata, N., Minobe, Y., Yano, M., and Sasaki, T. (1995). Conservation of duplicated segments between rice chromosomes 11 and 12. *Breed. Sci.* **45**, 373–376.
- Namuth, D.M., Lapitan, N.L.V., Gill, K.S., and Gill, B.S. (1994). Comparative RFLP mapping of *Hordeum vulgare* and *Triticum tauschii*. *Theor. Appl. Genet.* **89**, 865–872.
- Naranjo, T. (1995). Chromosome structure of *Triticum longissimum* relative to wheat. *Theor. Appl. Genet.* **91**, 105–109.
- Ogihara, Y., Liu, Y.G., Nagaki, K., Fujita, M., Kawaura, K., and Uozumi, M. (2000). Construction of large insert genomic DNA libraries of common wheat in a transformation-competent artificial chromosome (TAC) vector. In Eighth Plant and Animal Genome Conference, p. 84 (abstr).
- Panstruga, R., Buschges, R., and Schulze-Lefert, P. (1998). A contiguous 60 kb genomic stretch from barley provides molecular evidence for gene islands in monocot genomes. *Nucleic Acids Res.* **26**, 1056–1062.
- 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.
- Peng, J.R., 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.
- Qumsiyeh, M.B. (1994). Evolution of number and morphology of mammalian chromosomes. *J. Hered.* **85**, 455–465.
- Roberts, M.A., Reader, S.M., Dalglish, C., Miller, T.E., Foote, T., Fish, L.J., Snape, J.W., and Moore, G. (1999). Induction and characterization of *Ph1* wheat mutants. *Genetics* **153**, 1909–1918.
- SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake Berhan, A., Springer, P.S., Edwards, K.J., Avramova, Z., and Bennetzen, J.L. (1996). Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**, 765–768.
- SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., and Bennetzen, J.L. (1998). The paleontology of intergene retrotransposons of maize: Dating the strata. *Nature Genet.* **20**, 43–45.
- Sears, E.R. (1954). The aneuploids of common wheat. *Mo. Agric. Exp. Stn. Res. Bull.* **572**, 1–59.
- Tarchini, R., Biddle, P., Wineland, R., Tingey, S., and Rafalski, S. (2000). The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted co-linearity with maize chromosome 4. *Proc. Nat. Acad. Sci. USA*, in press.
- Theissen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J.T., Münster, T., Winer, K.-U., and Saedler, H. (2000). A short history of MADS-box genes in plants. *Plant Mol. Biol.* **42**, 115–149.
- Tranquilli, G., Lijavetzky, D., Muzzi, G., and Dubcovsky, J. (1999). Genetic and physical characterization of grain texture-related loci in diploid wheat. *Mol. Gen. Genet.* **262**, 846–850.
- 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.
- Wang, Z.M., Le Thierry d'Ennequin, M., Panaud, O., Gale, M.D., Sarr, A., and Devos, K.M. (2000). Trait mapping in foxtail millet. *Theor. Appl. Genet.*, in press.
- Wei, F., Gobelmann-Werner, K., Morroll, S.M., Kurth, J., Mao, L., Wing, R., Leister, D., Schulze-Lefert, P., and Wise, R.P. (1999). The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. *Genetics* **153**, 1929–1948.
- Zhang, H., Jia, J., Gale, M.D., and Devos, K.M. (1998). Relationship between the chromosomes of *Aegilops umbellulata* and wheat. *Theor. Appl. Genet.* **96**, 69–75.
- Zhang, H., Liu, X., Jia, J.Z., Gale, M.D., and Devos, K.M. (2000). Construction of a comparative genetic map between the genomes of wheat and *Aegilops longissima*. *Theor. Appl. Genet.*, in press.