Molecular Evolution of Receptor-Like Kinase Genes in Hexaploid Wheat. Independent Evolution of Orthologs after Polyploidization and Mechanisms of Local Rearrangements at Paralogous Loci[1]

Catherine Feuillet, Anja Penger, Klaus Gellner, Austin Mast, and Beat Keller*

Institutes of Plant Biology (C.F., B.K.) and Systematic Botany (A.M.), University of Zurich, CH–8008 Zurich, Switzerland; and Epidauros Biotechnologie AG, D–82347 Bernried, Germany (A.P., K.G.)

Hexaploid wheat is a young polyploid species and represents a good model to study mechanisms of gene evolution after polyploidization. Recent studies at the scale of the whole genome have suggested rapid genomic changes after polyploidization but so far the rearrangements that have occurred in terms of gene content and organization have not been analyzed at the microlevel in wheat. Here, we have isolated members of a receptor kinase (*Lrk*) gene family in hexaploid and diploid wheat, *Aegilops tauschii*, and barley (*Hordeum vulgare*). Phylogenetic analysis has allowed us to establish evolutionary relationships (orthology versus paralogy) between the different members of this gene family in wheat as well as with *Lrk* genes from barley. It also demonstrated that the sequences of the homoeologous *Lrk* genes evolved independently after polyploidization. In addition, we found evidence for gene loss during the evolution of wheat and barley. Analysis of large genomic fragments isolated from nonorthologous *Lrk* loci showed a high conservation of the gene content and gene organization at these loci on the homoeologous group 1 chromosomes of wheat and barley. Finally, sequence comparison of two paralogous fragments of chromosome 1B showed a large number of local events (sequence duplications, deletions, and insertions), which reveal rearrangements and mechanisms for genome enlargement at the microlevel.

Polyploidization has played a major role in higher plant evolution. A majority of the angiosperms (70%–80%; Masterson, 1994), including some of the most important crops (wheat, maize, potato, cotton, sugar cane), are polyploid. Polyploidization allows novel genetic interactions and its role in plant genome evolution is highly relevant (Wendel, 2000). Understanding the mechanisms underlying polyploid evolution can also have an impact on crop breeding, particularly for the development of new crop species such as Triticale (an allopolyploid of wheat and rye).

Bread wheat (*Triticum aestivum*) is hexaploid ($2n = 42$) with three (A, B, and D) subgenomes, each containing seven pairs of homoeologous chromosomes. Hexaploid wheat, which arose approximately 8,000 years ago (Feldman et al., 1995), is a classical example of allopolyploidization. It originated from the hybridization of three different diploid progenitors from the *Triticum* and *Aegilops* genera. The first step involved the hybridization between *Triticum urartu* Thum. Ex Gandil. (AA) and an unknown species (BB) related to *Aegilops speltoides* (SS). The resulting tetraploid wheat *Triticum turgidum* (AABB) then hybridized with *Aegilops tauschii* (DD) to produce the hexaploid bread wheat (AABBDD; Kihara, 1944; McFadden and Sears, 1946; Friebe and Gill, 1996). Compared with other allopolyploids, wheat is considered to be a young polyploid. The identity, the organization, and the evolution of the different genomes constituting wheat have been intensively studied in the last decades (for reviews, see Flavell et al., 1987; Kimber and Sears, 1987; Feldman et al., 1995). These studies were performed using a number of techniques such as cytogenetics, protein, and isozyme electrophoresis, comparative mapping and molecular markers, or DNA sequence comparisons. In addition, several tools that allow quick and efficient chromosomal localization in hexaploid wheat were developed, including a series of aneuploid lines (deletion, addition, or substitution lines) of the var Chinese Spring (Sears, 1966; Endo and Gill, 1996). These features, combined with the possibility of producing synthetic polyploids (Feldman et al., 1997), make wheat a model of choice to study the mechanisms underlying evolution in polyploid species.

Polyploidization events can have many consequences on genome evolution, particularly on gene expression and gene organization (for review, see Wendel, 2000). In wheat, recent studies (Feldman et al., 1997) with synthetic polyploids have indicated that genome reorganization probably occurs rapidly after the polyploidization event and that coding and non-coding regions might be differentially affected (Liu et al., 1998a, 1998b). So far, few studies have been performed to follow the fate (e.g. rate and type of changes) of individual loci after polyploid formation. A key question in studying gene evolution is

[1] This work was supported by the Swiss Priority Programme Biotechnology (grant nos. 5002–45033 and 5002–57824).

* Corresponding author; e-mail bkeller@botinst.unizh.ch; fax 41–16348204.

whether the genes have evolved independently or if there was a concerted evolution (Doyle and Gaut, 2000). Cronn et al. (1999) studied 16 individual loci corresponding to low copy sequences (including genes) in tetraploid cotton and its diploid progenitors. The authors found evidence for an independent evolution of these sequences after the formation of the tetraploid species. In hexaploid wheat, the power of low copy DNA sequence comparisons has not been well exploited due to the difficulty of cloning sequences from specific genomes and assessing the relationships (orthology versus paralogy) between the sequences. Paralogous genes arise by gene duplication, whereas orthologous genes arise by speciation (Fitch, 1970). In evolutionary studies, the distinction between the two cases can be difficult (Gogarten and Olendzenski, 1999). So far, no studies have analyzed the changes occurring between orthologous and paralogous sequences at the sequence and gene organization level in wheat.

Comparative genetics (Van Deynze et al., 1995) demonstrated that chromosome group 1 is very well conserved in the Triticeae. Moreover, cytogenetic (Gill et al., 1996) and microcolinearity studies (Feuillet and Keller, 1999) have indicated the presence of a conserved gene-rich region in the distal region of the short arm of chromosome group 1. This region is therefore a good target to study rearrangements at the microlevel. We have previously characterized a new family of receptor-like kinase genes (*Lrk*) in this region in the Triticeae (Feuillet et al., 1998; Keller and Feuillet, 2000). Six copies of *Lrk* genes were identified by Southern hybridization on the three homoeologous group 1 chromosomes in wheat. One copy was detected on chromosome 1A, three copies on 1B, and two copies on 1D (Feuillet et al., 1998).

Here, we analyzed the evolutionary relationships between all of the members of the *Lrk* gene family from the A, B, and D genomes of *T. aestivum*, the D genome of *A. tauschii*, the A^m genome of *Triticum monococcum* barley (which is related to the A genome donor of bread wheat, *T. urartu* (Dubcovsky et al., 1995) and the H genome of barley (*Hordeum vulgare*). We have found evidence for independent evolution of the genes after polyploidization that permitted us to assess the orthology of genes among the homoeologous chromosomes. In addition, sequence analysis of large genomic fragments (20–40 kb) of the *Lrk* loci from the 1B and 1D chromosomes has allowed us to compare the gene organization between non-orthologous regions and to determine at the microlevel the re-arrangements, which have occurred between recently duplicated genes.

RESULTS

Evolution of the *Lrk* Genes in Hexaploid Wheat, Diploid Wheat, *A. tauschii*, and Barley

To assess orthology among the *Lrk* genes on chromosome group 1 in the Triticeae, we reconstructed the evolutionary history of the homoeologous genes of hexaploid and diploid wheat, as well as sequences of *A. tauschii* and barley. To isolate all the members of the *Lrk* gene family from wheat, a fragment of 639-bp-encoding part of the extracellular domain of *Lrk* genes (Feuillet et al., 1998) was amplified by PCR from hexaploid wheat (var Chinese Spring) with two primers (G3/UP1) specific for the *Lrk* gene family on chromosome group 1 (Table I). Of 44 sequenced fragments, 10 individual sequences were identified. This was more than the six members we had previously identified by DNA hybridization (Feuillet et al., 1998), indicating that different genes might have resulted in the same restriction pattern on Southern blots. Specific primers were designed for each of the 10 *Lrk* genes (Table I) and were used to localize each gene on the homoeologous chromosomes in the aneuploid nulli-tetrasomic lines of var Chinese Spring

Table I. *Primer pairs used to amplify the extracellular domain of Lrk genes by PCR*

G3 and UP1 are located in very conserved regions of the *Lrk* genes and allow the amplification of all *Lrk* genes on chromosome group 1 in wheat. They can be used in combination with specific primers to amplify specifically one member of the gene family (*CSLrk*) and to determine the localization on the homoeologous chromosomes 1A, 1B, and 1D with aneuploid wheat lines. The annealing temperature (Ta) is indicated.

Specific Primer Sequence (5′ to 3′)	Ta	*Lrk* Gene	Chromosomal Location
G3 GAAAGATGAGTAAATTACTTG	55°C	*Lrk*	1A/1B/1D
UP1CTTGCAGTCCAACGGAAG			
14F1 ATGTCGCGAATCATGGAC	51°C	*CSLrk17*	1D
14R3 GTTTGTAGTCTTCGTACA			
G3/19R1 GTCCATTCGAAAAATTG	49°C	*CSLrk6*	1B
G3/25R1 GAATTCCATAGGTAAAG	49°C	*CSLrk9*	1B
G3/CS3 TGAGAATTGCCGATCACG	53°C	*CSLrk3*	1D
G3/CS16 GGGCCAGCAATGCTTTC	55°C	*CSLrk16*	1B
G3/CS11 GGGTTGCAGATGAACCCTC	55°C	*CSLrk11*	1B
G3/CS13 CAGAACTCCAAAGAAATATC	50°C	*CSLrk13*	1A
G3/CS5 GGTTGAAGATGAACTATC	48°C	*CSLrk5*	1A
G3/CS41 GAGTGCACTAGGGTTGCAGATGA	50°C	*CSLrk41*	?
UP1/CS22 CCTCCTGCTGCTACCTCTTAC	55°C	*CSLrk22*	1D

(Fig. 1). Two genes were located on chromosome 1A (*CSLrk5* and *CSLrk13*), four were found on 1B (*CSLrk6*, *CSLrk9*, *CSLrk11*, and *CSLrk16*), and three on 1D (*CSLrk3*, *CSLrk17*, and *CSLrk22*). One gene, *CSLrk41*, could not be localized. *Lrk* sequences from diploid bread wheat relatives were obtained from *A. tauschii* and from *T. monococcum.* Southern hybridization with pLrk10-A, encoding the extracellular domain of *Lrk10* (Feuillet et al., 1998), on genomic DNA from *A. tauschii* #309 and *T. monococcum* var DV92 showed four and two hybridizing fragments, respectively (data not shown). Here, four *Lrk* sequences were obtained by PCR amplification on *A. tauschii* genomic DNA with the G3/UP1primers (Table I). Two *T. monococcum Lrk* sequences were obtained from two bacterial artificial chromosome (BAC) clones isolated by screening the *T. monococcum* (DV92) BAC library (Lijavetzky et al., 1999) with pLrk10-A. The 10 *T. aestivum Lrk* DNA sequences, four *A. tauschii*, two *T. monococcum,* and two barley sequences (Feuillet and Keller, 1999) were compared in the phylogenetic analyses with a recently released *Lrk* sequence of oat (*rlk1a2*; AF237503) used as the outgroup. The optimal maximum parsimony (MP) phylogeny is shown in Figure 2A with the features shared with the maximum likelihood (ML) phylogeny indicated. The MP and ML searches each produced a single optimal tree. Combining the results of the chromosomal localization of the *Lrk* sequences with their phylogenetic relationships permits a partial description of orthology among the genes (Fig. 2B). Monophyletic lineages of genes are considered here to be orthologous if the lineage can be circumscribed to include not more than a single gene from each chromosome. The best supported examples of orthology (>70% bootstrap support) are between the A, B, and D genomes of *T. aestivum* and their diploid progenitors (Fig. 2A). The *CSLrk3*, *CSLrk17*, and *CSLrk22* genes from the D genome of *T. aestivum* are orthologous to the *TTLrk2*, *TT2Lrk2*, and *TTLrk1* genes from *A. tauschii*, respectively (Fig. 2B). The first pair of genes also appears to be orthologous to the *CSLrk16* gene of the B genome of *T. aestivum*. The *CSLrk5* gene from the A genome of *T. aestivum* is orthologous to the *TmLrk1* gene from *T. monococcum*, and the orthology of the unplaced *CSLrk41* gene with the *TmLrk2* gene of *T. monococcum* suggests that its position is likely in the A genome of *T. aestivum* (Fig. 2B). Orthologous relationships that are only weakly supported (<70% bootstrap) include those between: (a) *CSLrk5*, *TmLrk1*, *CSLrk9*, and *Hv1Lrk2*; (b) *CSLrk6* and *Hv1Lrk1*; and (c) *CSLrk22*, *TTLrk1*, *CSLrk41*, and *TmLrk2* (Fig. 2B). The position of the sequences of barley within lineages of genes from *Triticum* (Fig. 2A), rather than sister to the genes of *Triticum* and *Aegilops*, was unexpected. No orthologs from the D genome of *A. tauschii* and *T. aestivum* were found within these groups of orthology (Fig. 2B). This suggests either (a) a close phylogenetic relationship between barley and those genomes containing orthologous genes, or (b) extinction of orthologous genes in *A. tauschii* and the D genome of *T. aestivum* (c) mutations in the conserved primer binding sites. *CSLrk5* was found to be allelic with the *Lrk10* gene isolated from chromosome 1A in the hexaploid wheat variety Thatcher*Lr10* (Feuillet et al., 1997). *CSLrk5* appears to be orthologous to *Hv1Lrk2.* This has implications for the interpretation of the gene conservation previ-

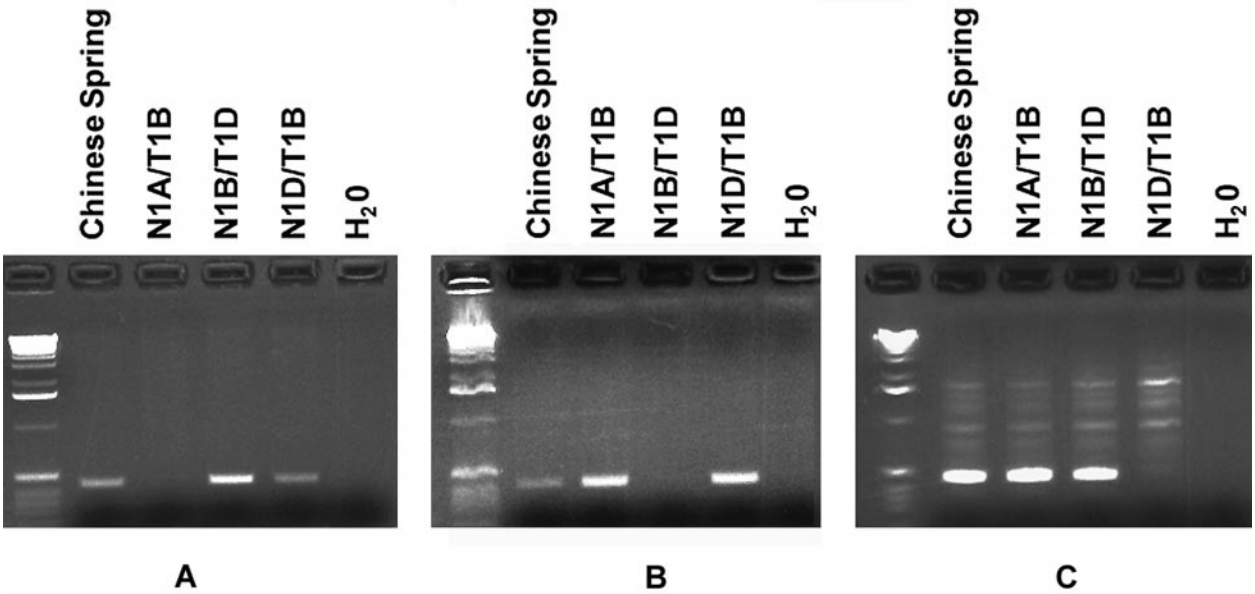

Figure 1. Chromosomal localization of *Lrk* genes on the group 1 chromosomes of hexaploid wheat. PCR was performed on genomic DNA of the var Chinese Spring and the derived aneuploid nulli-tetrasomic (N/T) lines of chromosome group 1 with primers specific for *CSLrk13* (A), *CSLrk9* (B), and *CSLrk17* (C). The absence of an amplified fragment indicated that *CSLrk13* is located on chromosome 1A, *CSLrk9* on chromosome 1B, and *CSLrk17* on chromosome 1D.

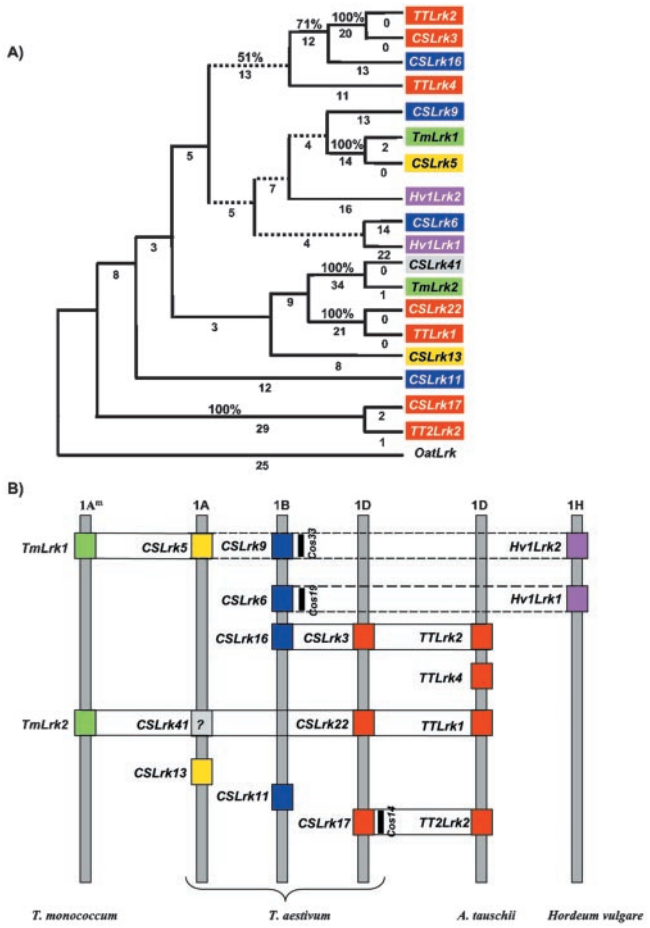

Figure 2. Evolutionary relationships between different *Lrk* gene family members in wheat, *A. tauschii*, and barley. Sequences of *T. aestivum* (var Chinese Spring), *T. monococcum* (var DV92), *A. tauschii* (no. 309), and barley (var Igri) were used. The colored boxes refer to the chromosomal location of each copy: green, $1A^m$ in *T. monococcum*; yellow, 1A in hexaploid wheat; blue, 1B in hexaploid wheat; red, 1D in hexaploid wheat and *A. tauschii*; and purple, 1H in barley. The sequence for which no chromosomal localization could be determined is indicated as a gray box. A, Phylogeny of the *Lrk* genes of hexaploid wheat, diploid wheat, *A. tauschii*, and barley species. The tree length of the MP tree was 331 steps (consistency index = 0.679); the −ln likelihood of the ML tree was 2,853.06 (data not shown). The sequence of a *Lrk* gene from oat was used as the outgroup to root the phylogenetic tree. Dotted lines indicate branches that are different between the MP and ML analysis. Bootstrap values >50% are indicated above the branches, and MP branch lengths are indicated below them. B, Schematic representation of orthologous and paralogous *Lrk* genes on the homoeologous chromosome 1 in *T. aestivum*, *A. tauschii*, *T. monococcum*, and barley. The solid lines indicate relationships shared with the ML analysis. The sequence *CSLrk41* for which a chromosomal localization could not be determined is placed on 1A according to the putative orthology to *TmLrk2*. The exact location of each gene on each chromosome and their relative positions are not known. The three cosmids (Cos19, Cos33, and Cos14) isolated from Thatcher*Lr10* are represented as bars along the allelic sequences from var Chinese Spring (*CSLrk9*, *CSLrk6*, and *CSLrk17*, respectively).

ously found at the *Lrk10* and *Hv1Lrk* loci (Feuillet and Keller, 1999).

The phylogenetic results also suggest that the paralogous *Lrk* genes have not undergone concerted evolution following polyploidization, which would have resulted in a topology that placed the paralogous sequences from single chromosomes together (Meagher et al., 1989). Pairwise distances between paralogs on each chromosome (7.7%–14.7%; computed for 535 aligned positions with ambiguously aligned regions excluded; data not shown), similarly, tended to be higher than pairwise distance values among orthologs (0.3%–14.4%; eight of the 17 pairwise comparisons were less than 7.7%).

Isolation and Coding Region Analysis of Three Cosmids Containing *Lrk* Genes from Chromosomes 1B and 1D in Hexaploid Wheat

To study the gene organization at *Lrk* loci on chromosomes 1B and 1D of hexaploid wheat, cosmids were isolated by screening a genomic library of the variety Thatcher*Lr10* with pLrk10-A, encoding the extracellular domain of *Lrk10* (Feuillet et al., 1998). Three positive cosmid clones (Cos14, Cos19, and Cos33) were selected and analyzed. On each of the three cosmids a gene homologous to *Lrk10* was identified (*Lrk14*, *Lrk19*, and *Lrk33*). The nucleotide sequences were compared with *Lrk10* and to the sequences obtained by PCR from the var Chinese Spring. This comparison clearly shows that *Lrk19*, *Lrk33*, and *Lrk14* are allelic to *CSLrk6*, *CSLrk9*, and *CSLrk17*, respectively. This indicates that Cos19 and Cos33 originated from chromosome 1B, whereas Cos14 originated from chromosome 1D. The phylogenetic analysis had shown that *CSLrk17* on chromosome 1D of *T. aestivum* is not orthologous to either *CSLrk9* or *CSLrk6* on chromosome 1B (Fig. 2B).

The three cosmids were completely sequenced (AF325196–98) with a coverage of 9, 8, and 11 times for Cos19 (35,872 bp), Cos33 (20,754 bp), and Cos14 (43,606 bp), respectively. Sequence analysis revealed the presence of putative coding regions. On the three cosmids, two receptor-like kinase genes of the *Lrk* and *Tak* type were identified (Feuillet and Keller, 1999). *Lrk19* (2,691 bp), *Lrk33* (2,908 bp), and *Lrk14* (3,378 bp) are found upstream of *Tak19–1* (2,322 bp), *Tak33* (2,352 bp), and *Tak14* (2,670 bp), respectively (Fig. 3, A–C). The number and position of the exons and introns are perfectly conserved with the *Lrk10* and *Tak10* genes on chromosome 1A (Fig. 3D; Feuillet and Keller, 1999). The length of the introns is more variable due to small insertions or deletions. The last 1,052 bp of Cos19 correspond to the end of a kinase domain of the *Tak* type suggesting the presence of a second *Tak* gene (*Tak19–2*) in this region on chromosome 1B (Fig. 3A). This sequence was different from the kinase domain of *Tak33* (Fig. 3B), suggesting that at least three *Tak* genes are present on chromosome

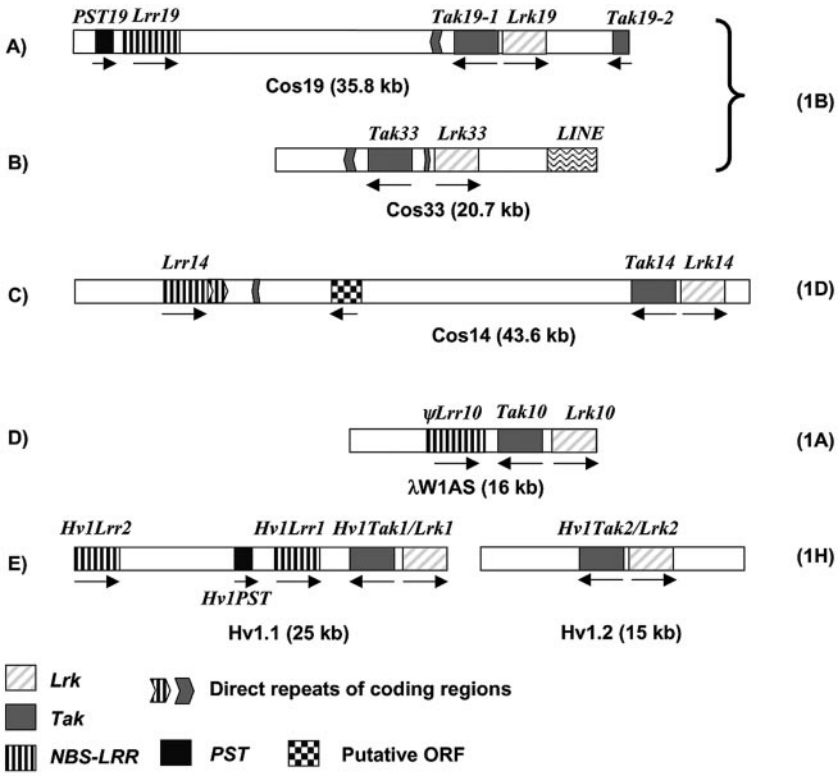

Figure 3. Comparison of the gene organization at the *Lrk/Tak* loci on genomic fragments originating from the 1A, 1B, and 1D chromosome of hexaploid wheat and from chromosome 1H of barley. The gene organization on three cosmid clones Cos19 on chromosome 1B (A), Cos33 on chromosome 1B (B), Cos14 on chromosome 1D (C) is compared with the gene organization on lambda clones from chromosome 1A (D) and chromosome 1H (E) previously described (Feuillet and Keller, 1999). Known genes and putative open reading frames are represented as indicated on the figure. The transcription orientation is indicated with arrows. Direct repeats are only indicated if they show homology to coding regions. The length of each genomic clone is given in kilobase pairs. The chromosomal localization is indicated on the right side of each clone.

1B. All *Lrk* and *Tak* genes are in a typical transcriptional opposite orientation as previously found on chromosome 1A and 1H (Feuillet and Keller, 1999). On Cos19 and Cos14, the length of the bidirectional promoter region is in the range of approximately 600 bp as found at the *Lrk10*, *Hv1Lrk1*, and *Hv1Lrk2* loci (Fig. 3, D and E). In contrast, the promoter region between *Lrk33* and *Tak33* (1,075 bp) contains an insertion of 539 bp corresponding to a direct repeat of sequences normally found at the 5′ end of *Tak* genes (Fig. 3B).

On Cos19 and Cos14, putative nucleotide binding site-Leu-rich repeat (NBS-LRR) genes were identified. *Lrr19* (2,760 bp), was located 18.7 kb downstream from the stop codon of *Tak19*, whereas *Lrr14* (2,781 bp) was found at 27 kb from *Tak14* (Fig. 3, A and C). Both coding regions showed more than 80% identity at the nucleotide level to *ψLrr10*, an NBS-LRR pseudogene that is located at 618 bp from *Tak10* on chromosome 1AS (Fig. 3D) (Feuillet and Keller, 1999). On both cosmids, the regions between the *Tak* and *Lrr* genes are larger than those observed at the *Lrk10* and *Hv1* loci (Fig. 3, A, C, D, and E). In the case of Cos19 this is clearly due to the insertion of transposable elements (see below). On Cos14, a sequence of 1,902 bp showing 50% similarity at the amino acid level to a predicted rice gene (AF111709) was detected in this region (Fig. 3C). It is interesting that this putative gene is located on rice chromosome 5, the homoeologous chromosome to the Triticeae chromosome group 1 (Van Deynze et al., 1995). No NBS-LRR gene was found on Cos33, which may be due to the shorter size of this cosmid. Finally, a putative coding region (PST19) similar to the predicted *Hv1PST* gene identified close to *Hv1Lrr1* (Feuillet and Keller, 1999) was detected 1,728 bp upstream from *Lrr19* (Fig. 3, A and E).

A number of direct repeats were identified on the three cosmids. Some of them corresponded to coding sequences. Hence, on each cosmid a direct repeat of the 3′ end of a *Tak* gene was detected. On Cos19 and Cos33 these repeats were found at approximately 2 kb downstream of *Tak19-1* and *Tak33*, respectively (Fig. 3, A and B), whereas on Cos14, the repeat was found at 24 kb from *Tak14* (Fig. 3C). A direct repeat of 814 bp (83% identity) of the LRR coding region of *Lrr14* was also found at 14 bp downstream from the stop codon. These data demonstrate that even at nonorthologous loci, the gene content and order at the *Lrk/Tak* loci is very well conserved between the 1A and 1B and 1D chromosomes in hexaploid wheat and on the barley chromosome 1HS.

Comparison of Paralogous *Lrk*/*Tak* Loci on Chromosome 1B

We have shown that Cos19 and Cos33 originated from chromosome 1B and carry *Lrk19* and *Lrk33*, which are allelic to the paralogous genes *CSLrk6* and *CSLrk9*, respectively. We can therefore assume that the genes on Cos33 are paralogous to the genes on Cos19, giving us a unique opportunity to study the evolution of such genes in wheat. Comparison of both cosmid sequences identified a region of 15.6 kb

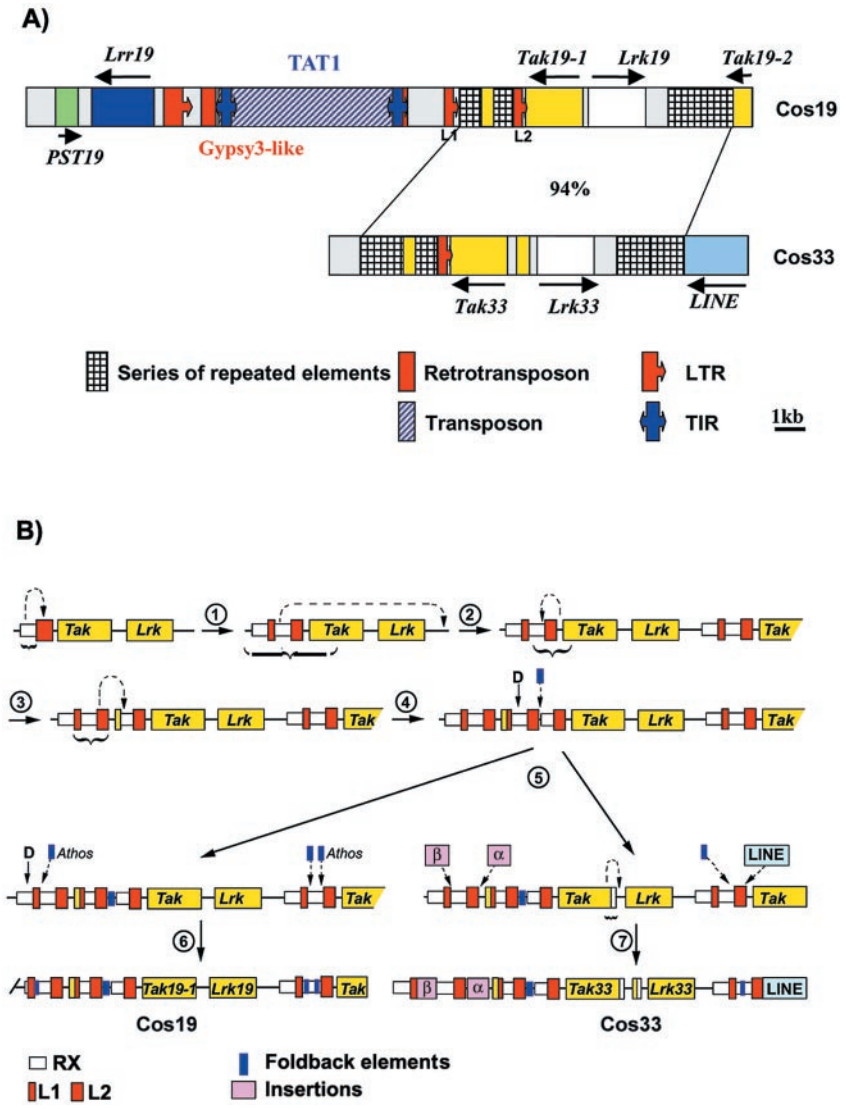

Figure 4. Gene organization and genome rearrangements at paralogous *Lrk* loci on chromosome 1B of hexaploid wheat. A, Detailed comparison of the gene organization on Cos19 and Cos33. Coding regions are represented according to the color code indicated on the figure with arrows indicating transcriptional orientation. Transposable elements (TAT1, LINE, and Gypsy retrotransposons) and their characteristic border elements (terminal inverted repeat [TIR] or long terminal repeat [LTR]) are indicated on both cosmids. The 15.6-kb conserved region (94% identity) is analyzed in more detail below. B, Model of the duplication events that have occurred in the flanking regions of the *Lrk/Tak* locus and have led to the duplication of the 15.6-kb region found on Cos19 and Cos33. 1 to 4, Duplication events before the duplication of the whole locus. The regions that have been duplicated are indicated with braces and the position of integration is depicted with the arrows. The exact length of the duplication at step 2 is not known. D indicates a short sequence deletion. 5, Duplication of a fragment of at least 15.6 kb on chromosome 1B. 6 and 7, Insertion of sequences (α and β) and elements (fold back elements, LINE retrotransposon) and deletions that have led to the actual sequence configuration of Cos19 and Cos 33.

containing the *Tak* and *Lrk* genes, which showed an overall identity of 94% (Fig. 4A). Coding regions were highly similar: *Tak19-1* and *Tak33* showed 93% identity, whereas *Lrk19* and *Lrk33* were 88% identical at the nucleotide level. High identity (94%–99%) was also found over several kilobases on both sides of the *Lrk* and *Tak* genes on Cos19 and Cos33. A detailed analysis of the non-coding regions revealed a very complex pattern of direct repeats, which have been successively inserted. A model representing the different steps leading to this pattern is given in Figure 4B and can be explained as follows: a sequence of approximately 1 kb (RX) was duplicated and inserted into the LTR of a Gypsy3 element, which was then split into two fragments, L1 and L2 (Fig. 4B-1). This new element plus 1 kb of the 3′ end of the *Tak* gene was then duplicated on the right side of *Lrk* (Fig. 4B-2). With the available sequence data we cannot determine the exact size of this duplication as it corresponds to the end of the cosmid sequences. Subsequently, two duplications containing different parts of the RX, L1/L2 sequences and 327 bp of the end of the *Tak* gene occurred, followed by a deletion of 100 bp and the insertion of a foldback element (Fig. 4, B-3 and B-4). At this stage, a fragment of at least 15.6 kb was duplicated on chromosome 1B (Fig. 4B-5). Following this duplication several events occurred independently on both copies to give rise to the two paralogous fragments found on Cos19 and Cos33. At least three foldback elements including two *Athos* elements (Wicker et al., our unpublished data) were inserted and a deletion of a RX sequence occurred on Cos19 (Fig. 4B-6). Two sequences of 1,440 bp (α) and 1,575 bp (β) delimited by a target site duplication of 15 bp were inserted into the repeats on Cos33 (Fig. 4B-7). A new type of fold back element and a duplication of 539 bp containing parts of *Tak19-1* and the promoter region were also inserted. Finally, a LINE retroelement integrated between the duplicate *Tak* sequence and L2 (Fig. 4B-7). These events are responsi-

ble for a sequence expansion of approximately 5 and 8 kb around the *Lrk/Tak* loci, respectively.

Analysis of the *Tak19-Lrr19* intergenic region (18.7 kb) of Cos19 showed the presence of an element of 7,184 bp, which we named TAT1 (Fig. 4A). TAT1 was delimited by a 6-bp target site duplication flanking two terminal inverted repeats (TIR) of 460 bp. The TIR contained a CACTA sequence that is characteristic of the CACTA type of transposons (Nacken et al., 1991). Within TAT1, a 1,742-bp sequence showed similarity to the TNP2 gene found in these transposons. Two LTRs of approximately 600 bp and partial sequences similar to the gag/pol precursors of a Gypsy3 retrotransposon (RT) flanked the TNP2/RT region (Fig. 4A), suggesting that the transposon TAT1 had inserted into a Gypsy3 retrotransposon. We conclude that a detailed analysis of paralogous sequences allowed us to identify small and local rearrangements, which occurred before and after the duplication and are responsible for a size increase of the analyzed region.

DISCUSSION

Paralogous and Orthologous *Lrk* Genes Have Evolved Independently and Have Been Differentially Deleted on the Homoeologous Triticeae Chromosomes

To study the pattern of evolution between the different homoeologous *Lrk* gene copies in wheat it was essential to characterize the phylogenetic relationships between all the different members of the gene family. A PCR-based strategy was used to isolate the complete family of *Lrk* genes present on chromosome group 1 in hexaploid wheat. The basis of the strategy was the definition of conserved regions from genes cloned earlier in wheat, barley, and *A. tauschii*. Although we cannot exclude that some members of the *Lrk* gene family (e.g. with mutations in the sequence of the conserved regions) are not represented here, we consider the set of genes analyzed in this study to be a good representation of this gene family. The correct identification of orthologous and paralogous relationships is essential in molecular comparative genetics as confusion between the two can lead to misinterpretations of the evolutionary mechanisms. Paralogy was established here with the use of aneuploid wheat lines (Sears, 1966). These lines offer a fast way to locate genes without the need of mapping each gene individually and have been extensively used to locate RFLP and sequence-tagged site markers in wheat (Sharp et al., 1989; Gill and Gill, 1994). In most cases we could differentiate between orthology and paralogy with good confidence based on the chromosomal localization of each copy and phylogenetic analysis of the sequence similarity. After polyploidization, orthologous genes can either continue to evolve independently or can undergo a concerted evolution. *Lrk* copies located on a specific chromosome in hexaploid wheat were phylogenetically sisters to the orthologous copies in the parental diploid species indicating an independent evolution of the *Lrk* genes since the polyploidization events. Greater pairwise distances between paralogous sequences than between orthologous ones lends additional support to this observation. Cronn et al. (1999) analyzed the relationships between 16 loci in diploid and tetraploid cotton species and also found evidence for independent evolution. However, few analyses based on sequence comparison have been made to study the evolution at specific loci after polyploidization in wheat (Talbert et al., 1998). It will be interesting to see whether different genes tell different stories about the evolution of coding regions after polyploidization.

Studies of synthetic polyploids in *Brassica*, cotton, and wheat have shown that polyploidization can lead to rapid and extensive genome changes including gene loss (Song et al., 1995; Feldman et al., 1997). Assessing the orthologous and paralogous relationships among the *Lrk* genes indicated that gene losses have occurred for this gene family in the Triticeae as well. The clearest examples of this are seen in the cases where orthologous *Lrk* genes to those in barley are absent on chromosome $1A^m$ of *T. monococcum*, 1A and 1D of *T. aestivum*, and 1D of *A. tauschii*. Previous phylogenetic analyses (Kellogg et al., 1996) have shown that *Hordeum* diverged first from the remaining Triticeae genomes. If a gene of *Hordeum* is found to be orthologous to a gene in any of the remaining genomes, then it is likely that an orthologous copy was present in their most recent common ancestor. Thus, assuming that we have a good representation of the entire gene family, each genome should have an orthologous gene by ancestry, unless gene loss has occurred.

Two of the *CSLrk* sequences, *CSLrk16* and *CSLrk41*, contained stop codons (data not shown) indicating that some members of the gene family are not expressed. Pseudogenization through accumulation of mutations is one of the possible consequences of the relaxation of selection pressure on a gene when several gene copies are present within a nucleus after polyploidization (Wendel, 2000). In hexaploid wheat, Galili and Feldman (1984) have found that in a gene family encoding endosperm proteins only the copies present on the B and D genome were expressed due to the silencing of the A genome homoeologs. We are currently investigating the expression pattern of the *Lrk* copies in the diploid genomes and whether the orthologous genes are expressed in polyploid wheat.

Conservation of Gene Organization and Reconstruction of the History of Duplicated *Lrk* Loci

In this study, we completely sequenced three cosmids clones containing genomic fragments of the *Lrk* region from chromosome 1B and 1D of *ThatcherLr10*.

The same gene organization (two receptor-like kinase genes and an NBS-LRR gene) was found on the cosmids although they originated from nonorthologous regions of chromosomes 1B and 1D. Preliminary analysis of the *Lrk* loci on BAC clones of *T. monococcum* and *A. tauschii* has confirmed this high conservation also in the diploid homoeologous A^m and D genomes (C. Feuillet and B. Keller, unpublished data). Thus, conservation of the gene organization that was observed at the *Lrk10* locus on chromosome 1A of Thatcher*Lr10* and on chromosome 1HS in barley (Feuillet and Keller, 1999) was not only due to the fact that we studied orthologous regions. This indicates that no major changes have occurred in this gene-rich region for at least 10 million years, when the barley and wheat lineage separated (Wolfe et al., 1989). A possible explanation for such conservation is that there is a selective advantage to keep the observed gene organization intact. One should also keep in mind that we studied a gene-rich region (one gene/5–10 kb; Feuillet and Keller, 1999; C. Feuillet and B. Keller, unpublished data from BACs), and we cannot exclude that the evolutionary mechanisms are different between gene-rich regions and regions containing more repetitive sequences.

The comparison of two paralogous loci on chromosome 1B (Cos19 and Cos33) has resulted in detailed insights into their molecular evolution. Whereas gene content and organization are well conserved, some of the non-coding regions show a great extent of reshuffling. Our model assumes a number of local duplications, deletions, and insertions, which can be assigned to the time before and after the locus duplication. Small duplications of sequences that have no feature of repetitive elements have successively occurred and have increased the size of the region by 5 to 8 kb. Thus, we are probably observing local mechanisms of genome size expansion that have to be considered in addition to those involving retrotransposons (Bennetzen, 2000). It is obvious that during further evolution, additional modifications of the region would result in a loss of detectable homology. This shows the power of comparing sequences of recent polyploids such as wheat to discover local mechanisms leading to rearrangements between the genomes. Further comparative studies of *Lrk* loci in species of *Triticum* and *Aegilops* with the B genome will determine the validity of our model of sequence rearrangements on chromosome 1B. With the increasing availability of genomic tools in wheat, libraries of *A. speltoides* (S genome, related to B) and *T. dicoccum* (AB genomes) should be available soon to allow this type of analysis.

A transposable element (TAT1) was detected on Cos19. TAT1 belongs to the CACTA type of transposons, which have been found in different plants species including grasses such as maize, rice, and sorghum (He et al., 2000). TAT1 is smaller than CACTA elements previously identified (15,164 bp for *Tam1* [Nacken et al., 1991]) and contains a single open reading frame similar to TNP2, which is thought to provide transposase activity. This suggests that TAT1 may be a new type of transposable element related to the CACTA type. To our knowledge, this is the first time that such a transposon is described in wheat. We are currently investigating the copy number of TAT1 in diploid and polyploid wheat species.

Among the grass species, wheat is a clear example of recent allopolyploidization. Large-scale sequencing and detailed analysis of the sequences will reveal more about the general mechanisms of wheat genome evolution. Such studies are of relevance to understand the genome of the most important crop plant as well as to reconstruct the history of genome development before and after polyploidization.

MATERIALS AND METHODS

Isolation and Chromosomal Localization of *Lrk* Genes from Hexaploid Wheat

Two primers (G3 and UPM1, Table I) were designed in two very conserved regions of the extracellular domain of *Lrk* genes (Feuillet et al., 1998). These primers were used to amplify fragments from hexaploid wheat (var Chinese Spring) and the diploid relatives *Aegilops tauschii* and *Triticum monococcum*. Fourty-four clones were sequenced from var Chinese Spring until all individual sequences were represented at least twice. Ten types of sequences were identified (GenBank accession nos. G67177 through 67185). Four types of sequences were amplified from *A. tauschii* (no. 309) (GenBank accession nos. G67176, G67187, G67188, and G67191). Two sequences were obtained by PCR amplification or subcloning fragments from *T. monococcum* (DV92) BACs (GenBank accession nos. G67186 and G67190). Sequencing was performed on an automatic DNA sequencer 4,200 (LI-COR, Lincoln, NE). Chromosomal localization of the sequences amplified from hexaploid wheat was made on a set of aneuploid nullisomic-tetrasomic lines of var Chinese Spring (Sears, 1966). For each gene a specific primer pair (Table I) was used in PCR amplification as previously described (Feuillet et al., 1997) with the annealing temperature indicated in Table I.

Isolation and Analysis of Cosmid Clones

A custom cosmid library was constructed from genomic DNA of the hexaploid wheat variety Thatcher*Lr10* (Stratagene, La Jolla, CA). Screening of the library was performed with ^{32}P-labeled pLrk10-A, encoding the extracellular domain of *Lrk10* (Feuillet et al., 1998). For sequencing, shotgun libraries were made from the three cosmids by shearing 10 to 20 μg of cosmid DNA (Hydroshear, Gene Machines, San Carlos, CA), blunt-ending the DNA fragments with mung bean exonuclease and size fractionation on an agarose gel. DNA fragments were dephosphorylated after incubation with calf intestine alkaline phosphatase (Boehringer Mannheim, Basel) and were subcloned into the

TOPO-Blunt end vector (Invitrogen, Carlsbad, CA). ElectroMAX DH10B *Escherichia coli* cells (Life Technologies, Basel) were then transformed by electroporation. Sequencing reactions were run on an ABI Prism 3,700 DNA (Perkin-Elmer Applied Biosystems, Foster City, CA) analyzer. Sequence gap closure was performed with internal primers either for PCR fragment generation and direct sequencing or internal sequencing of subclones. Sequence alignment was performed using Phrap Version 0.990329 and Consed Version 9.0 software. Detailed sequence analysis was performed with the Wisconsin Package Version 10.1, Genetics Computer Group (GCG; Madison, WI): sequence alignments were performed with the BestFit and PileUp programs, repeats were identified with the REPEAT, STEMLOOP, and BestFit programs, and coding regions were analyzed with the BLAST programs of GCG.

Phylogenetic Analysis

Alignment of the *Lrk* sequences was performed using PileUp from GCG and manually optimized to maintain the open reading frame. The highly variable regions (Feuillet et al., 1998) of the *Lrk* genes (aligned positions 183–207, 225–231, 425–494) were considered as ambiguously aligned and excluded from the phylogenetic analyses. The *Lrk* gene from oat, *rlk1a2* (AF237503), was used as an outgroup in all analyses because it is assumed to be an orthologous group 1 *Lrk* gene (K Armstrong, personal communication) from outside of the Triticeae tribe yet within the subfamily Pooideae. Phylogenetic hypotheses were generated using two alternative optimality criteria: maximum parsimony and maximum likelihood. The optimal MP and ML trees were found using heuristic searches that used Tree Bisection-Reconnection (TBR) branch swapping in PAUP (Phylogenetic Analysis Using Parsimony, version 4.0b4a, Sinaur Associates, Sunderland, MA). MP searches included 100 random addition replicates. The ML searches assumed a two-parameter model of nucleotide substitution (HKY model; Hasegawa et al., 1985) with among-site rate-heterogeneity assumed to be equal and a proportion of the sites assumed to be invariant among the sequences. All parameters within the model were estimated by the program during the ML search. The relative degree of branch support was determined within an MP framework using the bootstrap procedure (Felsenstein, 1985) in PAUP 4.0b4a. The original data set was resampled 100 times, and MP analyses of these replicates used TBR branch swapping.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Elena Conti for discussions regarding the phylogenetic analysis. We are also grateful to Nabila Yahiaoui, Christoph Ringli, and Nils Stein for critical reading of the manuscript. We would like to thank Geri Herren for his excellent technical assistance and Thomas Wicker for his help in designing the model of paralogous evolution.

Received December 7, 2000; accepted December 20, 2000.

LITERATURE CITED

Bennetzen JL (2000) Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell **12:** 1021–1029

Cronn RC, Small RL, Wendel JF (1999) Duplicated genes evolve independently after polyploid formation in cotton. Proc Natl Acad Sci USA **96:** 14406–14411

Doyle JF, Gaut BS (2000) Evolution of genes and taxa: a primer. Plant Mol Biol **42:** 1–23

Dubcovsky J, Luo MC, Dvorak J (1995) Differentiation between homoeologous chromosomes 1A of wheat and $1A^m$ of *Triticum monococcum* and its recognition by the wheat *Ph1* locus. Proc Natl Acad Sci USA **92:** 6645–6649

Endo TR, Gill BS (1996) The deletion stocks of common wheat. J Heredity **87:** 295–307

Feldman M, Liu B, Segal G, Abbo S, Levy AA, Vega JM (1997) Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. Genetics **147:** 1381–1387

Feldman M, Lupton FGH, Miller TE (1995) Wheats. *In* J Smartt, NW Simmonds, eds, Evolution of Crops, Ed 2. Longman Scientific, London, pp 184–192

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791

Feuillet C, 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

Feuillet C, Reuzeau C, Kjellbom P, Keller B (1998) Molecular characterization of a new type of receptor-like kinase (*wlrk*) gene family in wheat. Plant Mol Biol **37:** 943–953

Feuillet C, Schachermayr G, Keller B (1997) Molecular cloning of a new receptor-like kinase gene encoded at the *Lr10* disease resistance locus in wheat. Plant J **11:** 45–52

Fitch WM (1970) Distinguishing homologous from analogous proteins. Syst Zool **19:** 99–113

Flavell RB, Bennett MD, Seal AG, Hutchinson J (1987) Chromosome structure and organisation. *In* FGH Lupton, ed, Wheat Breeding, Its Scientific Basis. Chapman & Hall, London, pp 211–268

Friebe B, Gill BS (1996) Chromosome banding and genome analysis in diploid and cultivated polyploid wheats. *In* PP Jauhar, ed, Methods of Genome Analysis in Plants. CRC Press, New York, pp 39–60

Galili G, Feldman M (1984) Intergenomic suppression of endosperm protein genes in common wheat. Can J Genet Cytol **26:** 651–656

Gill KS, Gill BS (1994) Mapping in the realm of polyploidy: the wheat model. Bioessays **16:** 841–846

Gill KS, Gill BS, Endo TR, Taylor T (1996) Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. Genetics **144:** 1883–1891

Gogarten JP, Olendzenski L (1999) Orthologs, paralogs and genome comparisons. Curr Opin Genet Dev **9:** 630–636

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol **21:** 160–174

He ZH, Dong HT, Dong JX, Li DB, Ronald PC (2000) The rice *Rim2* transcript accumulates in response to *Magnaporthe grisea* and its predicted protein product shares similarity with TNP2-like proteins encoded by CACTA transposons. Mol Gen Genet **264:** 2–10

Keller B, Feuillet C (2000) Colinearity in the grass genomes. Trends Plant Sci **5:** 246–251

Kellogg EA, Appels R, Mason-Gamer RJ (1996) When genes tell different stories: the diploid genera of Triticeae (Gramineae). Syst Bot **21:** 321–347

Kihara H (1944) Discovery of the DD analyser, one of the ancestors of *T. vulgare.* Agric Hort **19:** 889–890

Kimber G, Sears ER (1987) Evolution in the genus *Triticum* and the origin of cultivated wheat. *In* EG Heyne, ed, Wheat and Wheat Improvement. American Society of Agronomy, Madison, WI, pp 154–164

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

Liu B, Vega JM, Feldman M (1998a) Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*: II. Changes in low-copy coding DNA sequences. Genome **41:** 535–542

Liu B, Vega JM, Segal G, Abbo S, Rodova M, Feldman M (1998b) Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*: I. Changes in low-copy noncoding DNA sequences. Genome **41:** 272–277

Masterson J (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science **264:** 421–424

McFadden ES, Sears ER (1946) The origin of *Triticum spelta* and its free-threshing hexaploid relatives. J Heredity **37:** 81–89, 107–116

Meagher RB, Berry-Lowe S, Rice K (1989) Molecular evolution of the small subunit of ribulose bisphosphate carboxylase: nucleotide substitution and gene conversion. Genetics **123:** 845–863

Nacken WKF, Piotrowiak R, Saedler H, Sommer H (1991) The transposable element *Tam1* from *Anthirrhinum majus* shows structural homology to the maize transposon En/Spm and has no sequence specificity of insertion. Mol Gen Genet **228:** 201–208

Sears ER (1966) Nullisomic-tetrasomic combinations in hexaploid wheat. *In* R Riley, KR Lewis, eds, Chromosome Manipulations and Plant Genetics. Oliver and Boyd, Edinburgh, pp 29–45

Sharp PJ, Chao S, Desai S, Gale MD (1989) The isolation characterization and application in the Triticeae of a set of wheat RFLP probes identifying each homoeologous chromosome arm. Theor Appl Genet **78:** 342–348

Song KM, Lu P, Tang K, Osborn TC (1995) Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc Natl Acad Sci USA **92:** 7719–7723

Talbert LE, Smith LY, Blake MK (1998) More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. Genome **41:** 402–407

Van Deynze AE, Dubcovsky J, Gill KS, Nelson JC, Sorrells ME, Dvorák J, Gill BS, Lagudah ES, McCouch SR, Appels R (1995) Molecular-genetic maps for group 1 chromosomes of Triticeae species and their relation to chromosomes in rice and oat. Genome **38:** 45–59

Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol **42:** 225–249

Wolfe KH, Gouy M, Yang Y-W, Sharp PM, Li W-H (1989) Data of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci USA **86:** 6201–6205