Funct Integr Genomics (2008) 8:149–164
DOI 10.1007/s10142-008-0073-z

ORIGINAL PAPER

Micro-colinearity between rice, *Brachypodium*, and *Triticum monococcum* at the wheat domestication locus *Q*

Justin D. Faris · Zengcui Zhang · John P. Fellers · Bikram S. Gill

Received: 23 November 2007 / Revised: 2 January 2008 / Accepted: 2 January 2008 / Published online: 22 January 2008

Abstract *Brachypodium*, a wild temperate grass with a small genome, was recently proposed as a new model organism for the large-genome grasses. In this study, we evaluated gene content and microcolinearity between diploid wheat (*Triticum monococcum*), *Brachypodium sylvaticum*, and rice at a local genomic region harboring the major wheat domestication gene *Q*. Gene density was much lower in *T. monococcum* (one per 41 kb) because of gene duplication and an abundance of transposable elements within intergenic regions as compared to *B. sylvaticum* (one per 14 kb) and rice (one per 10 kb). For the *Q* gene region, microcolinearity was more conserved between wheat and rice than between wheat and *Brachypodium* because *B. sylvaticum* contained two genes apparently not present within the orthologous regions of *T. monococcum* and rice. However, phylogenetic analysis of *Q* and leukotriene A-4 hydrolase-like gene orthologs, which were colinear among the three species, showed that *Brachypodium* is more closely related to wheat than rice, which agrees with previous studies. We conclude that *Brachypodium* will be a useful tool for gene discovery, comparative genomics, and the study of evolutionary relationships among the grasses but will not preclude the need to conduct large-scale genomics experiments in the Triticeae.

Keywords Wheat · Comparative genomics · Colinearity · Rice · Brachypodium

J. D. Faris (✉)
USDA-ARS Cereal Crops Unit,
Northern Crop Science Laboratory,
1307 18th Street North,
Fargo, ND 58105, USA
e-mail: justin.faris@ars.usda.gov

Z. Zhang
Department of Plant Sciences, Lofsgard Hall,
North Dakota State University,
Fargo, ND 58105, USA

J. P. Fellers
USDA-ARS Plant Science and Entomology Research Unit,
Throckmorton Plant Sciences Center, Kansas State University,
Manhattan, KS 66506, USA

B. S. Gill
Department of Plant Pathology,
Throckmorton Plant Sciences Center, Kansas State University,
Manhattan, KS 66506, USA

Introduction

The advent of molecular markers and molecular mapping of plant genomes has allowed researchers to conduct comparative mapping studies, which involve the comparison of the order and content of genes and molecular markers along chromosomes of related species. In the grasses, it has been shown that there is a high degree of genetic colinearity at the chromosome, or macro, level (Gale and Devos 1998; Devos and Gale 2000). For example, early studies comparing marker colinearity among members of the Poaceae including wheat (*Triticum aestivum* L.), rice (*Oryza sativa* L.), barley (*Hordeum vulgare* L.), rye (*Secale cereale* L.), oats (*Avena sativa* L.), and maize (*Zea mays* L.) revealed remarkable similarities in macrocolinearity (Ahn et al. 1993;

Van Deynze et al. 1995a,b). Whereas the genomes of closely related species such as wheat and barley were shown to be colinear across most of the genome, the genomes of the more distantly related grasses such as rice, maize, and oats could be divided into homologous linkage blocks that corresponded to segments of the wheat and barley genomes.

The degree of macrocolinearity observed between rice, which has a compact genome, and other grasses with large genomes such as wheat and barley led to the notion that rice could be used as a vehicle for gene discovery and genomic analysis of the large-genome grasses. The availability of the completed rice genome sequence (International Rice Genome Sequencing Project 2005) made it possible to make comparisons of colinearity at the deoxyribonucleic acid (DNA), or micro, level between the rice genome and local sequences of wheat or barley. Some such studies reported good levels of microcolinearity between wheat and rice (Yan et al. 2003; Chantret et al. 2004; Distelfeld et al. 2004; Schnurbusch et al. 2007) but most have indicated multiple breaks in microcolinearity because of inversions, deletions, duplications, and other rearrangements (Bennetzen 2000; Feuillet and Keller 2002; Li and Gill 2002; Sorrells et al. 2003; Francki et al. 2004; Lagudah et al. 2006; Lu and Faris 2006; Valarik et al. 2006; Bossolini et al. 2007). These results indicate that it is necessary to exercise extreme caution when utilizing the rice genome information for map-based cloning of genes in wheat.

The limited degree of colinearity observed in wheat–rice comparative studies is reflective of the fact that rice diverged from the Triticeae lineage approximately 50 million years ago (Mya) (Paterson et al. 2004). Recently, *Brachypodium distachyon* and *B. sylvaticum* have been proposed as new model organisms to study genomics of large-genome cereals (Draper et al. 2001). *Brachypodium* is a wild temperate grass with many characteristics ideal for a model organism. For example, it has one of the smallest genome sizes among the grasses at approximately twice the size of *Arabidopsis* (Bennett and Leitch 2005; Vogel et al. 2006a). In addition, the plants are physically small, have simple growth requirements, are inbreeding by nature, and have a rapid generation time. *Brachypodium* is amenable to transformation by *Agrobacterium* and biolistics (Vogel et al. 2006a), and genomic resources such as expressed sequence tags (ESTs) have become available (Vogel et al. 2006b). Efforts are under way to sequence the *Brachypodium* genome, which will allow extensive comparative analysis with rice and other grasses.

Phylogenetic analysis based on internal transcribed spacer regions and the chloroplast *ndh*F gene (Catalan and Olmstead 2000), restriction fragment length polymorphism (RFLP) and random amplification of polymorphic DNA markers (Catalan et al. 1995), and EST sequences from 20 highly expressed genes (Vogel et al. 2006b) all indicate that *Brachypodium* is more closely related to wheat and barley than is rice. Based on the comparative analysis of 23 expressed genes, Bossolini et al. (2007) estimated that wheat and *Brachypodium* diverged about 35–40 Mya, compared to 50 Mya for the divergence of the Triticeae lineage from rice (Paterson et al. 2004).

Indeed, the data collected to date suggests that *Brachypodium* may be a better model organism for the large-genome grasses than is rice. However, very few studies have been conducted to evaluate the level of conserved microcolinearity between wheat and *Brachypodium* to shed light on the potential utility of *Brachypodium* as a surrogate for gene discovery and genome analysis. Bossolini et al. (2007) evaluated a 371-kb region in *B. sylvaticum* and compared it with orthologous regions of wheat and rice. *Brachypodium* and wheat had perfect colinearity among genetic markers, whereas rice possessed a 220-kb inversion in the region studied, indicating that *Brachypodium* was more conserved with wheat at the microlevel. However, *Brachypodium* and wheat were found to differ considerably in gene content within the region indicating that *Brachypodium* would not be a perfect surrogate for genomic studies in wheat.

Additional comparative analysis experiments and molecular data are needed to determine the extent of microcolinearity between *Brachypodium* and wheat. Bossolini et al. (2007) evaluated colinearity at a locus containing the major leaf rust resistance gene *Lr34*. Disease resistance loci may be subjected to strong selection pressure because of the need for them to evolve rapidly as corresponding pathogens acquire virulence. On the other hand, genes that are involved in inflorescence and flower development such as the domestication gene *Q* tend to be highly conserved. Therefore, one might hypothesize that the colinearity of genes near floral homeotic loci would be more conserved than those embedded near disease resistance loci. The *Q* gene in wheat is considered a predominant domestication gene responsible for the widespread cultivation of bread wheat (*T. aestivum*, $2n=6x=42$, AABBDD genomes) and durum (*T. turgidum*, $2n=4x=28$, AABB genomes) because it confers the free-threshing (naked grain) phenotype while also pleiotropically affecting spike morphology, rachis fragility, glume tenacity, plant height, and days to heading (Leighty and Boshnakian 1921; Mackey 1954; Muramatsu 1963; 1986; Kato et al. 1999; Faris and Gill 2002; Faris et al. 2003; Simons et al. 2006). We isolated the *Q* gene by map-based cloning and found that it had a high degree of similarity to members of the AP2 family of plant transcription factors (Simons et al. 2006). In this study, we report the evaluation of microcolinearity between a genomic region of diploid wheat (*T. monococcum* L., $2n=2x=14$, A^mA^m genomes) containing the *Q* locus and orthologous regions of *B. sylvaticum* and rice.

Materials and methods

BAC identification and sequencing

The *T. monococcum* BAC contig of approximately 290 kb in size spanning the *Q* locus on chromosome 5A was described in Faris et al. (2003). BACs 598P15, 448N4 (combined GenBank no. AY914083), and 594O11 (GenBank no. AY914082) were completely sequenced and assembled as described in Faris et al. (2003). BAC 122I22 (GenBank no. AY914080) was also described in Faris et al. (2003) but was sequenced with lower coverage than the others resulting in more than 30 contigs. Fingerprinting and pulsed-field gel electrophoresis (PFGE) of this clone indicated that it is approximately 80 kb in size and overlaps with 50 kb of 594O11 (Faris et al. 2003). Although we are not confident in the order of 122I22 contigs that do not overlap with 594O11, we included it in this study because one nonoverlapping contig contained a gene found within a colinear region of rice and *Brachypodium* (see "Results").

The *B. sylvaticum* BAC library consisting of 30,228 clones representing 6.6 genome equivalents (Foote et al. 2004) was obtained from the John Innes Centre, UK. The Tmap2 probe (GenBank no. AY170867; Faris et al. 2003), which is a fragment of the *Q* gene, was radiolabeled and hybridized to the high-density filters of the *B. sylvaticum* bacterial artificial chromosome (BAC) library as described in Faris et al. (2000). DNA from positive BAC clones was isolated, and the approximate sizes of the clones were determined using PFGE as described in Faris et al. (2003). The positive BACs were fingerprinted by digesting them with *Hin*dIII followed by separation of the restriction fragments in 0.9% agarose gels. BAC DNA from each clone was spotted onto nylon membranes and hybridized with fragments of the *T. monococcum* leukotriene A-4 hydrolase (*TmLH*) and 40S ribosomal S23 (*Tm40S*) genes, which were both known to exist on the *T. monococcum* BAC contig, to identify a *B. sylvaticum* BAC containing putative orthologs of all three genes (*BsLH*, *Bs40S*, and *BsQ*). The *B. sylvaticum* BAC 23D12 (GenBank no. EU153459) was found to be positive for all three genes and was selected for sequencing. Shotgun library construction, sequencing, and sequence assembly were done by the Washington Genome Sequencing Center, Washington University, St. Louis, MO. The orthologous 50-kb segment of rice chromosome 3 corresponding to position 34,270,000–34,320,000 was downloaded from Gramene (http://www.gramene.org/; re et al. 2002).

BAC annotation

The *T. monococcum*, *B. sylvaticum*, and rice sequences were analyzed using TEnest (http://www.plantgdb.org/prj/TE_nest/TE_nest.html) with the appropriate repeat sequence databases to annotate the sequences for transposable elements (TEs) and to obtain sequence files with masking of the identified TEs. Masked sequences were then subjected to BLASTx (Altschul et al. 1997) searches of the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/BLAST/) to reveal additional putative TEs previously unidentified ($<e^{-30}$), which were then masked manually. Masked sequences were submitted to the Rice Genome Automated Annotation System (http://www.//ricegaas.dna.affrc.go.jp/), where sequences were analyzed with the various integrated coding prediction programs (Autopredgenset, GENSCAN, RiceHMM, FGENESH), homology search analysis programs (BLAST, HMMER, ProfileScan, MOTIF), and repetitive DNA analysis programs (RepeatMasker, Printrepeats, AutoPredLTR, BLASTn against NCBI LTRdb, and BLASTx against NCBI transposon subjects). Genes were considered candidates only if at least two ab initio gene finders predicted open reading frames (ORFs).

For reasons outlined in Bossolini et al. (2007), we adopted their criteria for distinguishing real genes from those that are likely to contain parts of TEs to allow a conservative estimate of the number of genes present in each of the three species. All colinear genes were considered to be real genes, but all noncolinear genes were subjected to additional scrutiny. First, predicted genes found within TE boundaries such as long terminal repeats (LTRs) or terminal inverted repeats were considered parts of TEs. Second, predicted genes were subjected to BLASTx searches against all *Arabidopsis* proteins at http://www.arabidopsis.org/. Predicted genes that did not have significant ($<e^{-8}$) homology to *Arabidopsis* proteins were also considered as parts of TEs. As Bossolini et al. (2007) mentioned, this scrutiny runs the risk of eliminating genes that are specific to monocots, but combined with the indication that they are not colinear, the risk is rather small.

The *Q* gene full-length complementary DNA (cDNA) sequence (GenBank no. AY645945) was aligned with genomic sequences of the *AP2*-like orthologs from *B. sylvaticum* and rice to determine correct intron/exon boundaries. All other gene sequences were subjected to BLASTn searches of the NCBI dbEST database (http://www.ncbi.nlm.nih.gov/dbEST/) and the Institute for Genomic Research gene indices (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/Blast/index.cgi) to identify corresponding ESTs and/or tentative consensus (TC) sequences, respectively. EST/TC sequences with at least 80% identity were used to annotate exon/intron splice junctions of the genes. The coding sequences (CDS) were then used as queries in BLASTx searches of the NCBI non-redundant database to identify homologs and assign putative functions using a relatively liberal threshold of $<e^{-8}$ and at least 30% amino acid identity.

Comparative sequence analysis

B. sylvaticum and *T. monococcum* coding sequences were used as queries in BLASTn and tBLASTx searches of rice genomic sequences using Gramene (http://www.gramene.org/; Ware et al. 2002) with significance thresholds of less than e^{-20} with greater than 65% identity and less than e^{-10} with greater than 30% identity, respectively. *B. sylvaticum* gene sequences were tested for similarities to the *T. monococcum* BAC contig, and all rice genes were tested for similarities to the *T. monococcum* and *B. sylvaticum* sequences using bl2seq (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi; Tatusova and Madden 1999). In addition, all noncoding sequences from each species were tested for similarities to the sequences of the others to identify conserved noncoding sequences or putative conserved coding sequences that might have been missed by the ab initio gene finders.

Amino acid alignments were done using ClustalW as part of the MacVector v8.0 software package. Phylogenetic trees were constructed from ClustalW alignments of the genomic DNA or amino acid sequences using the unweighted pair group method with arithmetic mean (UPGMA) and multiple distance-based methods available in the MacVector v8.0 software. Confidence values for nodes were calculated using 1,000 bootstraps.

Molecular mapping

For *B. sylvaticum* genes that were not colinear with *T. monococcum*, polymerase chain reaction (PCR) primers were either designed based on the *B. sylvaticum* gene sequences and amplified from BAC 23D12 or based on homologous wheat ESTs and used to amplify the corresponding sequences from cDNA of the *T. aestivum* cultivar Chinese Spring (CS) using standard PCR protocols. The PCR products were separated on 2.0% agarose gels and gel purified using the Wizard SV Gel and PCR Clean Up System (Promega, Madison, WI). Purified fragments were used as probes and hybridized to Southern blots containing the CS nullisomic–tetrasomic (NT) lines (Sears 1954), in which an extra pair of homologous chromosomes compensates for the absence of a pair of chromosomes. Probes that detected fragments on chromosome 5A were then hybridized to Southern blots containing CS and the CS–*T. dicoccoides* chromosome 5A disomic substitution line (CS-DIC 5A), which are the parents of the F_2 population using for positional cloning of the *Q* locus (Faris et al. 2003). The two parents were digested with enzymes *Apa*I, *Bam*HI, *Bgl*II, *Dra*I, *Eco*RI, *Eco*RV, *Hin*dIII, *Sac*I, *Sca*I, and *Xba*I. Polymorphic fragments were mapped in 178 F_2 plants using the enzyme giving the clearest polymorphism. The markers were placed relative to existing markers on the linkage map of CS×CS-DIC 5A using the "TRY" command and a logarithm of the odds threshold of 3.0 in Mapmaker v.2.0 for Macintosh (Lander et al. 1987).

Results

Gene content of the *T. monococcum*, *B. sylvaticum*, and rice sequences

Masking of the 260,380-bp *T. monococcum* BAC sequence using TEnest and manual means resulted in the masking of 154,866 bp or 60% of the total sequence. Initial analysis of the masked sequence indicated that it contained 13 predicted ORFs. However, seven of these ORFs were eliminated as putative genes because they were either within typical TE boundaries, contained portions of TEs or did not have corresponding *Arabidopsis* homologs. The 30 kb of BAC 122I22 contigs that did not overlap with the 260-kb contig contained one gene (*Tm40S*). Therefore, a total of seven *T. monococcum* genes were identified (Table 1, Fig. 1).

Among the seven *T. monococcum* genes, *Tm40S*, the *Q* gene ortholog (*TmQ*), and a putative leukotriene A-4 hydrolase protein (*TmLH*) were strongly supported by wheat ESTs and had significant similarities to known proteins allowing us to ascribe putative functions (Table 1). The full-length genomic sequence for *Tm40S* was present in a 1,056-bp contig from BAC 122I22. Alignments of the *T. monococcum* genomic sequence containing the 595-bp *Tm40S* gene revealed strong hits to a large number of wheat ESTs, the best being CJ856456 with 98% identity. *Tm40S* contains three exons and has a predicted protein of 136 amino acids. *TmQ* is 3,533 bp in length with ten exons, a predicted protein of 447 amino acids, and occupies position 55,308–58,840 of the *T. monococcum* contig (Table 1, Fig. 1). The *T. aestivum* and *T. monococcum Q* genes share 98% identity at both the coding sequence and amino acid levels. The *TmLH* gene is 3,521 bp in size and resides at position 244,761–247,678 of the *T. monococcum* BAC contig (Fig. 1). The gene has four exons and encodes a predicted protein of 613 amino acids. *TmLH* had 98% identity at the DNA level to a *T. aestivum* EST (BU100661).

The remaining four *T. monococcum* genes were paralogs coding for hypothetical proteins with F-box domains (*TmFB1–TmFB4*). Although no *Arabidopsis* proteins with homology to the *TmFB* genes were identified, the *TmFB* genes were not eliminated because a homologous gene was found in a colinear region of *B. sylvaticum* (see below). The *TmFB* genes were supported by wheat ESTs but had no significant similarity to any known proteins in the databases (Table 1). It is most likely that not all the *TmFB* genes are functional. Two exons were predicted for *TmFB1*, which is

Table 1 Characteristics of genes within the *Triticum monococcum* BAC contig and their best EST, protein, and rice genomic sequence hits

Gene	Position	Size (bp)	Number of exons	Number of amino acids	EST hit	*e* value	NCBI BLASTx hit	*e* value	Rice BLASTn			Rice tBLASTx		
									Chromosome	Position	*e* value	Chromosome	Position	*e* value
Tm40S	?	595	3	136	CJ856456 *Triticum aestivum* cDNA	1*e*−114	NP_001051708 40S ribosomal protein S23 (*Oryza sativa*)	6*e*−65	3	34,286,900–34,287,062	5*e*−134		34,286,487–34,287,060	1*e*−60
TmQ	55,308–58,840	3,533	10	447	DY61160 *Triticum aestivum* cDNA	0.0	AY170867 Floral homeotic protein (Q) (*Triticum monococcum*)	0.0	3	34,297,692–34,300,922	1*e*−209	3	34,297,433–34,300,125	1*e*−10
TmFB1	87,169–91,912	4,744	2	617	CJ546490 *Triticum aestivum* cDNA	0.0	–	–	–	–	–	–	–	–
TmFB2	104,532–105,878	1,347	4	449	CJ654347 *Triticum aestivum* cDNA	0.0	–	–	–	–	–	–	–	–
TmFB3	110,664–112,022	1,359	1	452	CJ546490 *Triticum aestivum* cDNA	3*e*−162	–	–	–	–	–	–	–	–
TmLH	244,761–248,282	3,521	4	613	BU100671 *Triticum aestivum* cDNA	0.0	EAY92357 leukotriene A-4 hydrolase (*Oryza sativa*)	6*e*−93	3	34,317,821–34,321,063	0.0	3	34,317,821–34,321,064	6*e*−204
TmFB4	255,299–256,510	1,113	1	370	CJ654347 *Triticum aestivum* cDNA	3*e*−92	–	–	–	–	–	–	–	–

4,744 bp at position 87,169–91,912 of the contig (Table 1). Closer evaluation of the gene revealed that *TmFB1* contains a 2,891-bp insertion at position 1,279 of the gene with 92% similarity to the gypsy LTR retrotransposon Sukkula_643D12-2, which is likely to disrupt its function (Fig. 2). The insertion of the retroelement occurred after the duplications that gave rise to *TmFB2*, *TmFB3*, and *TmFB4* because it is not present in those copies. In addition, the stop codon of *TmFB1* is disrupted, which results in a larger predicted ORF and amino acid sequence (617) compared to the paralogous *TmFB* genes. The predicted genomic sequence of *TmFB2* is 1,347 bp at position 104,532–105,878 of the contig. Alignments of this gene with the other *TmFB* genes indicated that it contains a 33-bp deletion at the 5′ end, which includes the start codon predicted for the other *TmFB* genes. *TmFB2* also contains internal deletions of 44, 241, 62, and 13 bp (Fig. 2). *TmFB3* and *TmFB4* are 1,359 and 1,113 bp in size and reside at positions 110,664–112,022 and 255,332–256,510, respectively (Table 1, Fig. 1). *TmFB3* contains a full-length coding sequence and is most likely to be functional producing a protein consisting of 452 amino acids. *TmFB4* has a mutation that induces a premature stop codon at position 1,113 of the gene. *TmFB1*, *TmFB2*, and *TmFB3* are in close proximity and in the same orientation, whereas *TmFB4* resides at the end of the contig approximately 143 kb from *TmFB3* and is in the opposite orientation (Fig. 1). These data suggest that *TmFB1*, *TmFB2*, and *TmFB4* are pseudogenes, whereas *TmFB3* may be a functional gene.

The average gene density within the *T. monococcum* contig was one gene per 41 kb, but the genes were not dispersed evenly. A distance of 133 kb between *TmFB3* and *TmLH* contains no genes, whereas *TmQ*, *TmFB1*, *TmFB2*, and *TmFB3* all lay within a 57-kb segment (Fig. 1).

Screening of the *B. sylvaticum* BAC library with the TmAP2 probe revealed six positive clones. Analysis of the *Hin*dIII fingerprints showed that all six clones shared many common fragments indicating that all were derived from a common region of the *B. sylvaticum* genome (data not shown). Dot blot analysis indicated that BAC 23D12 contained sequences homologous to *Tm40S* and *TmLH*, as well as *TmQ*, and thus was selected for sequencing.

A total of 16,550 bp (17%) of the 97,698-bp *B. sylvaticum* BAC 23D12 were masked by TEnest and manual means. Initial analysis indicated there were 11 predicted ORFs. Four of these ORFs, which were not colinear with either *T. monococcum* or rice, were eliminated as candidates because they either contained portions of TEs or were not homologous to any *Arabidopsis* protein. Therefore, the *B. sylvaticum* BAC contained seven putative genes (Table 2). The first gene on the BAC (*BsMIIP*) resides at position 602–938 bp and had slight similarity to the wheat EST CA735993, which corresponds to the wheat TC sequence TC266461. Alignments between TC266461 and the *BsMIIP*-

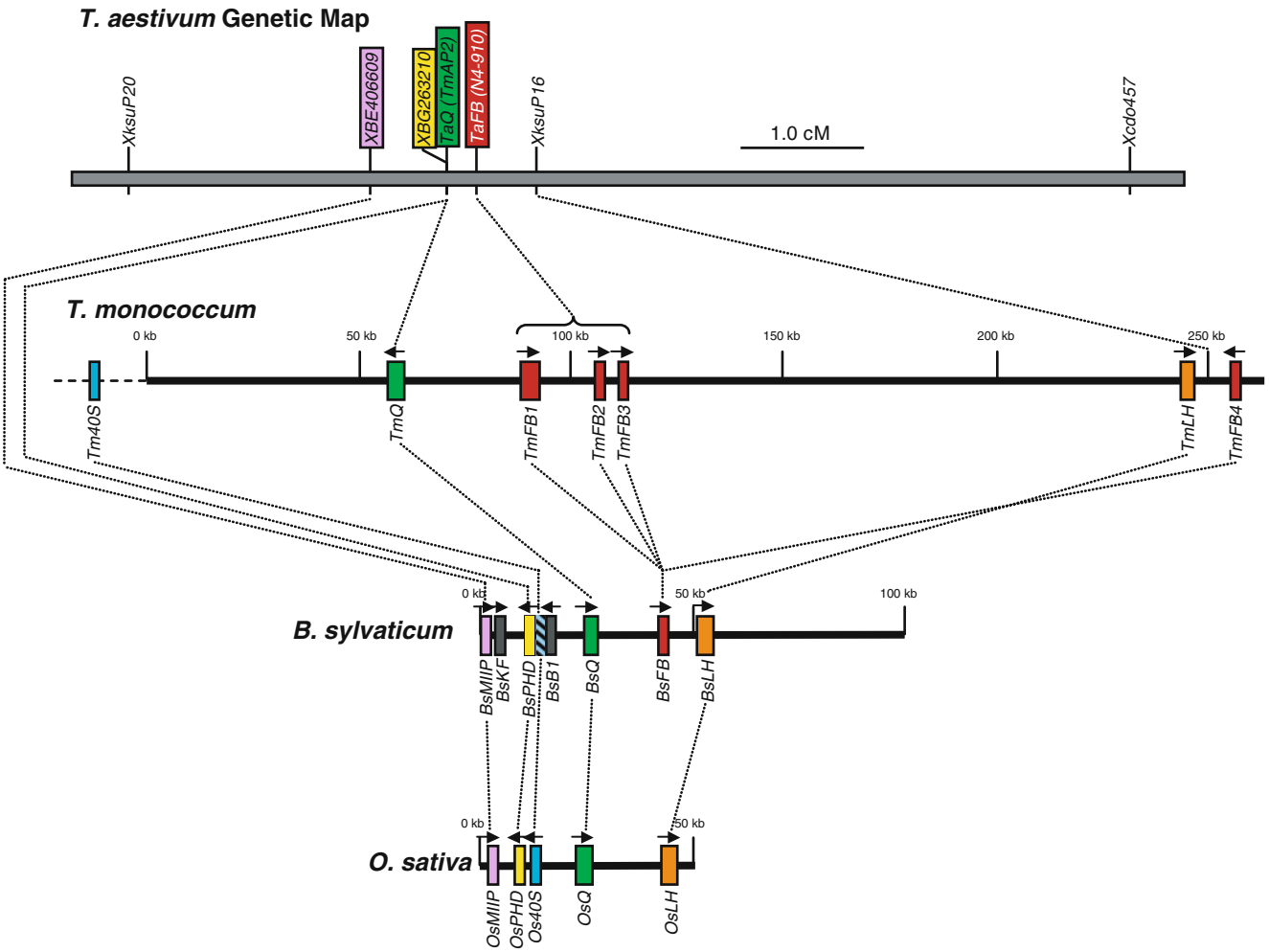


Fig. 1 Microcolinearity between the *Q* locus of *Triticum monococcum*, *Brachypodium sylvaticum*, and rice. Genes are shown as *colored boxes* along the physical maps of each species, and transcriptional orientations are indicated by *arrows above the boxes*. A kilobase (*kb*) scale is shown above each physical map. The *black*- and *blue-hatched box* on the *B. sylvaticum* map indicates a degenerate gene. Orthologous genes are connected by *dotted lines*. The *T. aestivum* genetic map of the chromosome 5A *Q* region derived from CS×CS-DIC 5A (Faris et al. 2003) is shown at the *top* and was used to determine the genetic locations of the EST-based markers *XBE406609* and *XBG263210*, which are orthologous to *BsMIIP*/*OsMIIP* and *BsPHD*/*OsPHD*, respectively

predicted coding region (data not shown) indicated that only the 3′ end of the gene consisting of the last two exons was present on the BAC, and the 5′ end was missing. BLASTx searches of the database using the *BsMIIP*-predicted coding region indicated only slight similarity to a multiple inositol polyphosphate phosphatase (MIIP) protein (Table 2), but the same search using the corresponding wheat TC sequence (TC266461) as a query showed very strong hits (e=0.0) to a *T. aestivum* MIIP PhyIIa1 protein (DQ995971). Together, these results indicate that a portion of a MIIP-like gene resides at the very beginning of the *B. sylvaticum* BAC clone 23D12.

The second gene (*BsKF*) on the *B. sylvaticum* BAC clone was supported by a *B. distachyon* root EST and resides at position 4,035–6,129 bp, has four exons, and codes for a putative protein consisting of 182 amino acids, which had significant similarity to a Keltch repeat-containing F-box protein from rice (Table 2). The third *B. sylvaticum* gene (*BsPHD*) at position 13,202–16,398 was also supported by a *B. distachyon* EST. *BsPHD* has five exons and a predicted protein of 265 amino acids with similarity to a PHD finger domain-containing protein from rice (Table 2). *BsB1* contains two exons and occupies position 19,041–20,698 bp. This gene was supported by a wheat EST (BJ239406) and has a predicted protein of 85 amino acids with similarity to a rice BRICK1-like protein.

The *B. sylvaticum Q* gene ortholog (*BsQ*) is at position 25,299–28,545 on BAC 23D12 (Table 2). *BsQ* has ten exons and a predicted protein of 445 amino acids. Alignments of the full-length *T. aestivum Q* cDNA with *BsQ* showed that they shared 80% identity at the nucleic acid level (data not shown).

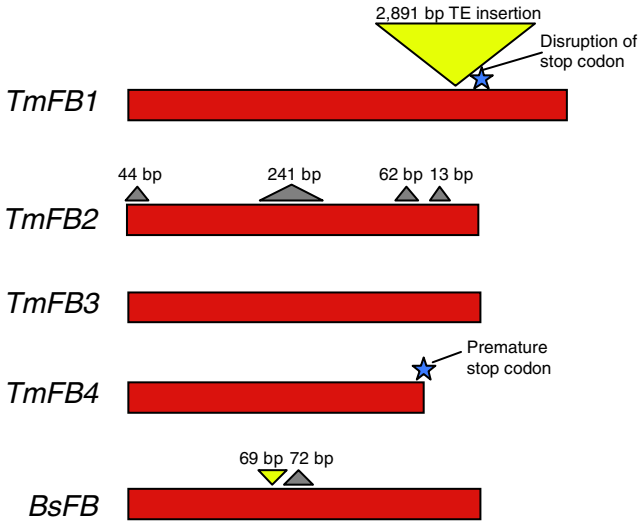


Fig. 2 Structural comparison of the four paralogous F-box genes in *Triticum monococcum* (*TmFB1–TmFB4*) and the *Brachypodium sylvaticum* ortholog *BsFB*. *TmFB3* has a full-length open reading frame, whereas the others have structural anomalies. *Yellow* and *gray triangles* represent insertions and deletions, respectively. *Blue stars* represent the interruption or induction of a stop codon

BsFB occupies position 45,199–46,549 bp of the BAC and was predicted to contain one exon and a protein of 449 amino acids (Table 2). *BsFB* was supported by a *Saccharum officinarum* cDNA (AY645945) but had no significant homology to any sequences in the database when used as a query in BLASTx searches.

The last gene (*BsLH*) on the *B. sylvaticum* BAC, at position 51,185–55,297 bp, has four exons and codes for a predicted protein of 272 amino acids (Table 2). *BsLH* was well supported by a rice cDNA clone and significant BLASTx hits to leukotriene A-4 hydrolase-like proteins. However, compared to other highly conserved *LH* genes, *BsLH* contains a 1,976-bp insertion at position 2,524 of the genomic sequence, which had similarity to a hypothetical gene (GenBank EAZ16708) on rice chromosome 10. The insertion was intronic and does not appear to alter the predicted protein (Fig. 3).

The presence of seven genes within the 97 kb *B. sylvaticum* BAC gives an average density of one gene per 14 kb, but as in *T. monococcum*, the genes were not evenly distributed (Fig. 1). All seven genes lied within the first 55 kb of the BAC giving an estimated density of one gene per 8 kb, whereas the remaining 42 kb contained no genes.

Analysis of the 50-kb region extending from 34,270,000 to 34,320,000 bp of rice chromosome 3 using TEnest resulted in the masking of 6,485 bp (13%). Initial analysis of genes within the orthologous region of rice chromosome 3 indicated the presence of nine predicted ORFs. However, we eliminated four of them from the analysis because they either possessed portions of putative TEs or had no homo-

Table 2 Characteristics of genes within the 97 kb *Brachypodium sylvaticum* BAC and their best EST, protein, and rice genomic sequence hits

Gene	Position	Size (bp)	Number of exons	Number of amino acids	EST hit	*e* value	NCBI BLASTx hit	*e* value	Rice BLASTn			Rice tBLASTx		
									Chrosome	Position	*e* value	Chromosome	Position	*e* value
BsMIIP	602–938	?	2+	?	CA735993 *Triticum aestivum* cDNA clone	7*e*−11	NP_001051705 Multiple inositol polyphosphate phosphatase (*Oryza sativa*)	9*e*−08	-	-	-	3	34,278,322–34,278,629	1*e*−10
BsKF	4,035–6,129	2,095	4	182	DV486613 *Brachypodium distachyon* root EST	3*e*−30	NP_001067606 Keltch repeat-containing F-box (*Oryza sativa*)	2*e*−27	-	-	-	11	7,874,009–7,874,625	4*e*−20
BsPHD	13,202–16,398	3,197	5	265	DV472714 *Brachypodium distachyon* callus EST	0.0	NP_001051707 PHD finger domain containing (*Oryza sativa*)	5*e*−97	3	34,282,461–34,285,350	7*e*−120	3	34,282,437–34,285,562	6*e*−57
BsB1	19,041–20,698	1,658	2	85	BJ239406 *Triticum aestivum* cDNA	1*e*−102	NP_001048610 BRICK1 (*Oryza sativa*)	1*e*−36	2	35,674,136–35,675,440	2*e*−81	2	35,674,136–35,675,461	3*e*−26
BsQ	25,299–28,545	3,247	10	445	AY645945 *Triticum aestivum* floral homeotic protein (Q) mRNA	0.0	AAU94924 Floral homeotic protein (*Triticum urartu*)	2*e*−92	3	34,297,708–34,300,938	1*e*−213	3	34,297,705–34,301,011	7*e*−80
BsFB	45,199–46,549	1,350	1	449	CA125561 *Saccharum officinarum* cDNA	6*e*−60	–	–	–	–	–	–	–	–
BsLH	51,185–55,297	4,113	4	272	AK073744 *Oryza sativa* cDNA	0.0	NP_001051710 leukotriene A-4 hydrolase (*Oryza sativa*)	1*e*−149	3	34,317,284–34,321,060	0.0	3	34,317,221–34,321,064	8*e*−286

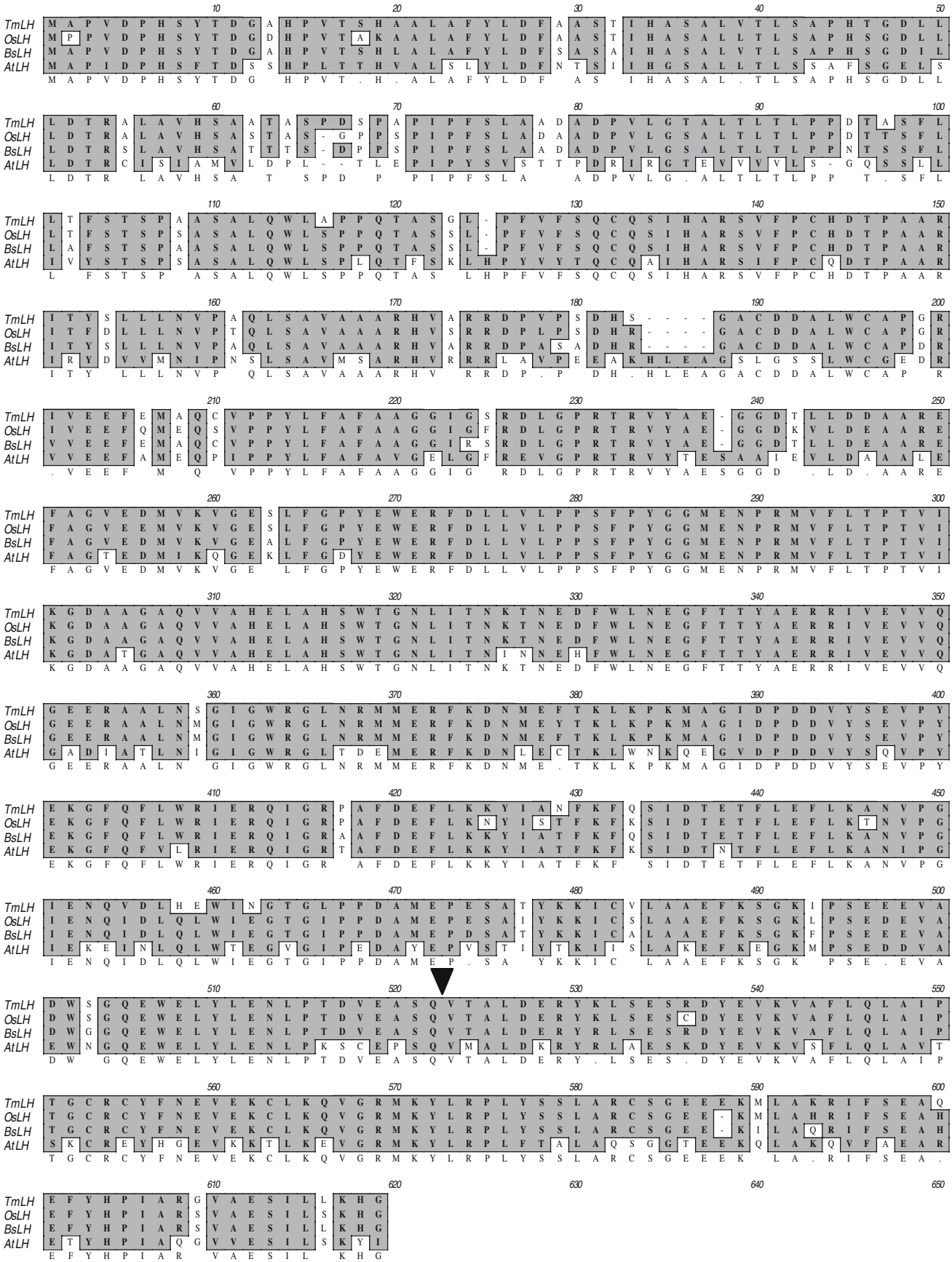

TmLH
OsLH
BsLH
AtLH

Fig. 3 Alignment of deduced amino acid sequences of the leukotriene A-4 hydrolase (*LH*)-like genes from *Triticum monococcum* (*TmLH*), rice (*OsLH*), *Brachypodium sylvaticum* (*BsLH*), and *Arabidopsis* (*AtLH*). The *black triangle* indicates the position of a 1,976-bp insertion in the *BsLH* gene. The consensus sequence is shown at the *bottom of the alignment*

logs in *Arabidopsis*. Therefore, the 50-kb rice chromosome 3 segment under evaluation contained five genes (Table 3). BLASTx alignments revealed putative functions of all of the genes, which included a MIIP-like gene (*OsMIIP*), a PHD finger domain-containing protein (*OsPHD*), a 40S ribosomal protein S23 (*Os40S*), the *Q* gene ortholog (*OsQ*), and a leukotriene A-4 hydrolase like protein (*OsLH*). The average gene density of this region in rice is one gene per 10 kb, and all five genes were distributed more or less evenly within the 50 kb (Fig. 1).

Microcolinearity of *T. monococcum*, *B. sylvaticum*, and rice at the *Q* locus

BLASTn and tBLASTx searches of the rice genome using the *B. sylvaticum* gene *BsMIIP* as a query revealed no significant BLASTn hits, but a significant similarity to *OsMIIP* on rice chromosome 3 was identified using tBLASTx (Table 2). *BsMIIP* and *OsMIIP* lie in the same orientation and are likely orthologous, but no corresponding *MIIP*-like ortholog was identified on the *T. monococcum* BAC contig (Fig. 1). The wheat EST CA735993, which had homology to *BsMIIP*, is contained within the 3′ end of wheat TC266461. We found that the 5′ end of TC266461 contained the wheat EST BE406609, which has been physically mapped by the National Science Foundation (NSF) wheat EST project to the long arms of wheat homoeologous group 5 chromosomes (http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi query: BE406609). We used BE406609 as an RFLP probe and mapped it in the CS×CS-DIC 5A F_2 population to determine its location on the genetic map relative to previously mapped markers, including the *Q* locus. Marker *XBE406609* mapped 0.7 cM proximal to the *Q* locus indicating that the *MIIP*-like genes in wheat, *B. sylvaticum*, and rice were colinear (Fig. 1).

No sequences homologous to *BsKF* were identified in the *T. monococcum* BAC contig, and searches of the rice genome revealed significant tBLASTx hits to a region of rice chromosome 11 (Table 2). A wheat EST (BG313181) with significant homology to *BsKF* was identified and found to have been mapped by the NSF wheat EST project to the long arms of homoeologous group 2 chromosomes (http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi query: BG313181). This provides strong evidence that *BsKF* is not colinear with *T. monococcum* and rice, and the apparent interruption of colinearity by *BsKF* may have resulted from a transposition event after the divergence of *Brachypodium* and the Triticeae (Fig. 4). However, we cannot rule out the possibility that another homolog of *BsKF* other than the one mapping to wheat group 2 chromosomes exists within the colinear region of wheat chromosome 5A.

Sequences with homology to *BsPHD* were not present in the *T. monococcum* BAC contig, but queries of the rice genome indicated an orthologous gene (*OsPHD*) in the same orientation on chromosome 3 (Table 2, Fig. 1). Searches of the Triticeae ESTs using *BsPHD* as a query revealed significant similarity to the wheat EST BG263210, which was mapped to the long arms of homoeologous group 5 chromosomes by the NSF wheat EST project (http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi query: BG263210). Molecular mapping of BG263210 as an RFLP probe in the CS×CS-DIC 5A population indicated that it cosegregated with the *Q* gene (Fig. 1). Therefore, the *PHD* finger-like genes from wheat, *B. sylvaticum*, and rice are colinear.

Putative 40S ribosomal genes were identified in *T. monococcum* (*Tm40S*) and rice (*Os40S*) and were colinear. However, no 40S ribosomal-like gene was identified in the *B. sylvaticum* BAC sequence, suggesting that an ortholog was not present. Alignments of *Tm40S* and *Os40S* with the *B. sylvaticum* 23D12 BAC sequence revealed a segment of about 300 bp at position 17,136–17,439 bp with 86 and 83% identity, respectively. No ORFs were predicted at this position indicating that only remnants of the orthologous 40S ribosomal gene are present in this region of *B. sylvaticum*. However, the position of this homologous

Table 3 Rice genes within the 50-kb segment of chromosome 3 orthologous to the *Q* locus

Gene	Reference identification	Position (chromosome 3)	Putative function
OsMIIP	NP_001051705	34,273,673–34,279,067	Multiple inositol polyphosphate phosphatase
OsPHD	NP_001051707	34,281,847–34,285,679	PHD finger protein
Os40S	NP_001051708	34,286,109–34,288,349	40S ribosomal protein S23
OsQ	Q84TB5_ORYSJ	34,297,664–34,301,257	Floral homeotic protein
OsLH	NP_001051710	34,317,091–34,321,295	Leukotriene A4 hydrolase

Reference identifications, positions, and assigned functions were taken from the Gramene website (http://www.gramene.org/).

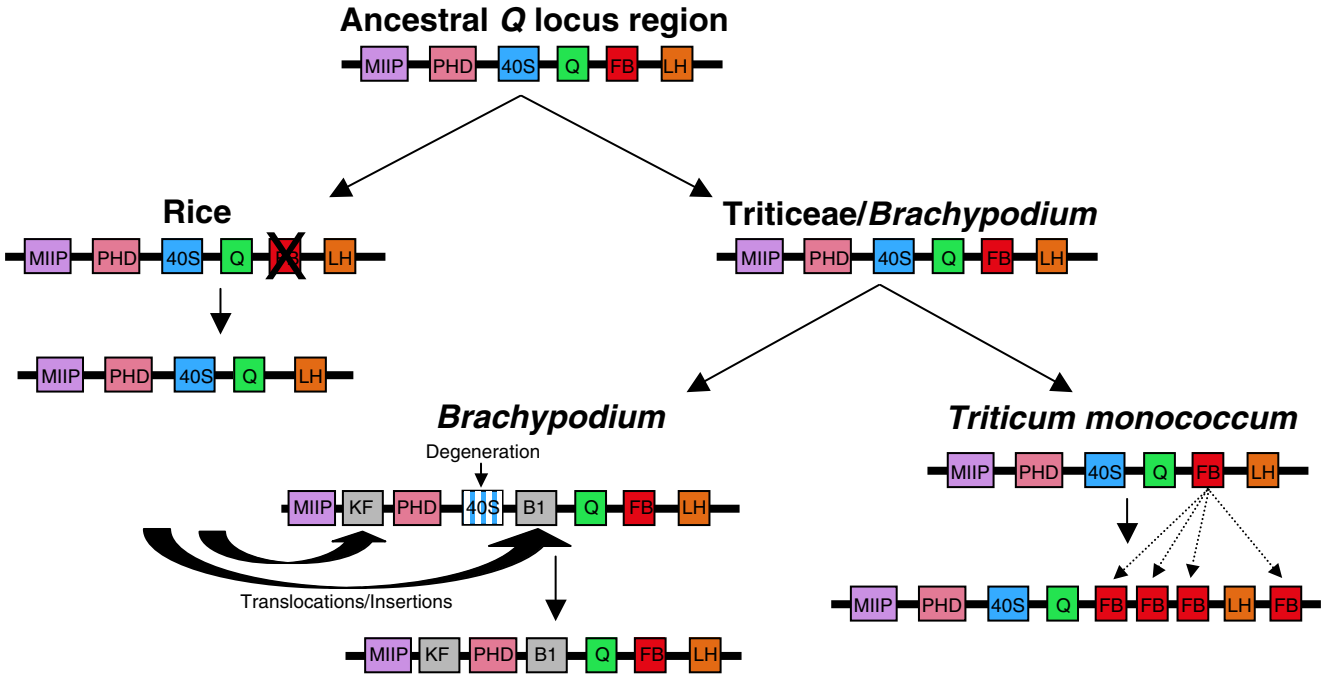


Fig. 4 Model of the evolution of the *Q* locus in rice, *Brachypodium*, and *Triticum monococcum*. The hypothetical ancestor locus contains one copy of each of the six colinear genes. A *crossed-out* gene indicates a deletion event. *Hatched* gene indicates degeneration. *Large black arrows* indicate transposition/insertion events, and *dotted arrows* indicate duplication events

sequence in *B. sylvaticum* retains colinearity with *Tm40S* and *Os40s* (Fig. 1).

Analysis of the *BsB1* gene revealed no significant similarities with the *T. monococcum* BAC contig, but a putative ortholog was identified on rice chromosome 2. We used a fragment of the *BsB1* gene to probe the wheat NT lines, but no hybridization occurred (data not shown). The wheat EST BJ239406 with significant homology to *BsB1* was PCR amplified and hybridized to the wheat NT digested with multiple enzymes, but no intergenomic polymorphism was observed, and therefore, the chromosomal location of the EST could not be determined. As with *BsKF*, the presence of *BsB1* in this region of the *Brachypodium* genome suggests it occurred by transposition after the divergence of Triticeae and *Brachypodium* (Fig. 4).

Q gene orthologs were obviously present in *T. monococcum*, *B. sylvaticum*, and rice because that was the criteria we used to define the regions under investigation. Although *TmQ*, *BsQ*, and *OsQ* were colinear, *TmQ* was in the opposite orientation compared to *BsQ* and *OsQ* (Fig. 1).

No sequences with homology to *BsFB* were identified in rice using either BLASTn or tBLASTx searches, which was not unexpected given that no significant similarities were identified in BLASTx searches of the protein database (Table 2). Alignment of *BsFB* with the *T. monococcum* contig revealed that it was homologous to all four *T. monococcum* F-box genes (data not shown). *TmFB1*, *TmFB2*, and *TmFB3* are colinear with *BsFB* and in the same orientation, whereas *TmFB4* is not colinear and in the opposite orientation suggesting that *TmFB4* was recently translocated or involved in an inversion (Fig. 1). The fact that only one F-box gene is present in the *B. sylvaticum* BAC suggests that duplication of the F-box genes in *T. monococcum* took place after the divergence of the Triticeae and *Brachypodium* (Fig. 4).

TmLH and *OsLH*, which are orthologs of the *BsLH* gene, were present in the *T. monococcum* contig and rice chromosome 3, respectively (Fig. 1). The three genes were in the same orientation and in a colinear arrangement.

Comparison of intergenic distances

The intergenic distances of colinear genes between *B. sylvaticum* and rice were generally larger in *B. sylvaticum* (Fig. 1). *BsMIIP* and *BsPHD* were separated by 12,264 bp in *B. sylvaticum*, whereas *OsMIIP* and *OsPHD* were separated by 2,780 bp in rice. This is a sixfold increase in physical distance between *BsMIIP* and *BsPHD* in *B. sylvaticum*, which includes the presence of the *BsKF* gene. Physical distance between *Q* and *PHD* orthologs were more similar. *BsQ* and *BsPHD* were 8,901 bp apart in *B. sylvaticum*, and *OsQ* and *OsPHD* were 11,985 bp apart in rice. This intergenic distance included the presence of *BsB1* and the remnants of a 40S ribosomal-like gene in *B. sylvaticum* and the *Os40S* gene in rice. The physical

distance between *BsLH* and *BsQ* in *B. sylvaticum*, which includes the *BsFB* gene, was 22,640 bp. The distance between the corresponding rice orthologs *OsQ* and *OsLH* was 15,834 bp. Therefore, only the distance between orthologous *PHD* and *Q* genes was larger in rice compared to *B. sylvaticum*, whereas the distances among other colinear orthologs were slightly larger in *B. sylvaticum*.

Intergenic distances among colinear genes were greater in *T. monococcum* in all cases compared to *B. sylvaticum* and rice (Fig. 1). *Tm40S* and *TmQ* were separated by at least 55,308 bp, whereas there is only 9,315 bp separating *Os40S* and *OsQ* in rice. The actual *TmFB* ortholog of *BsFB* is not known, but *TmQ* is 28,329 bp from *TmFB1*, which is the closest of the four paralogs, whereas *BsQ* is 16,654 bp from *BsFB* in *B. sylvaticum*. Because rice does not contain a gene orthologous to the F-box-like genes, we also compared the distances between *Q* and *LH* orthologs among the three species. *TmQ* and *TmLH* are 185,921 bp apart, compared to the distances of 22,640 and 15,834 bp in *B. sylvaticum* and rice, respectively. The genome expansion observed in *T. monococcum* is primarily due to the presence of TEs and to a lesser degree the duplication of the *TmFB* genes.

Comparative sequence analysis of colinear genes

Complete sequences of the orthologous *Q* and *LH* genes were available from all three species as well as *Arabidopsis* and were used to conduct phylogenetic analysis based on predicted amino acid sequences. For evaluation of the *Q* gene orthologs, we also included orthologs of *T. aestivum*, barley, and the maize *ids1* gene (Chuck et al. 1998). ClustalW alignments indicated the amino acid sequences among the seven species are highly conserved within the AP2-DNA binding domains (Fig. 5). The *Arabidopsis* AP2 protein (*AtQ(AP2)*) is extremely divergent in the 5′ and 3′ regions flanking the AP2 domains compared to the orthologous grass genes. The first 28 amino acids are quite conserved among the grasses, as are several additional domains within the region upstream of the first AP2 domain. The 3′ ends of the proteins downstream of the second AP2 domain are also quite conserved among the grasses with a couple exceptions. The *Brachypodium* ortholog has an apparent insertion at position 426, and maize and rice have insertions at position 476 (Fig. 5). Percent overall amino acid identities ranged from 42% for *AtQ(AP2)* compared to *TaQ*, *TmQ*, *BsQ*, and *OsQ,* to 98% for *TaQ* and *TmQ* (Table 4). Phylogenetic analysis based on the orthologous *Q* sequences indicated that *B. sylvaticum* is more closely related to members of the Triticeae than are maize and rice (Fig. 6).

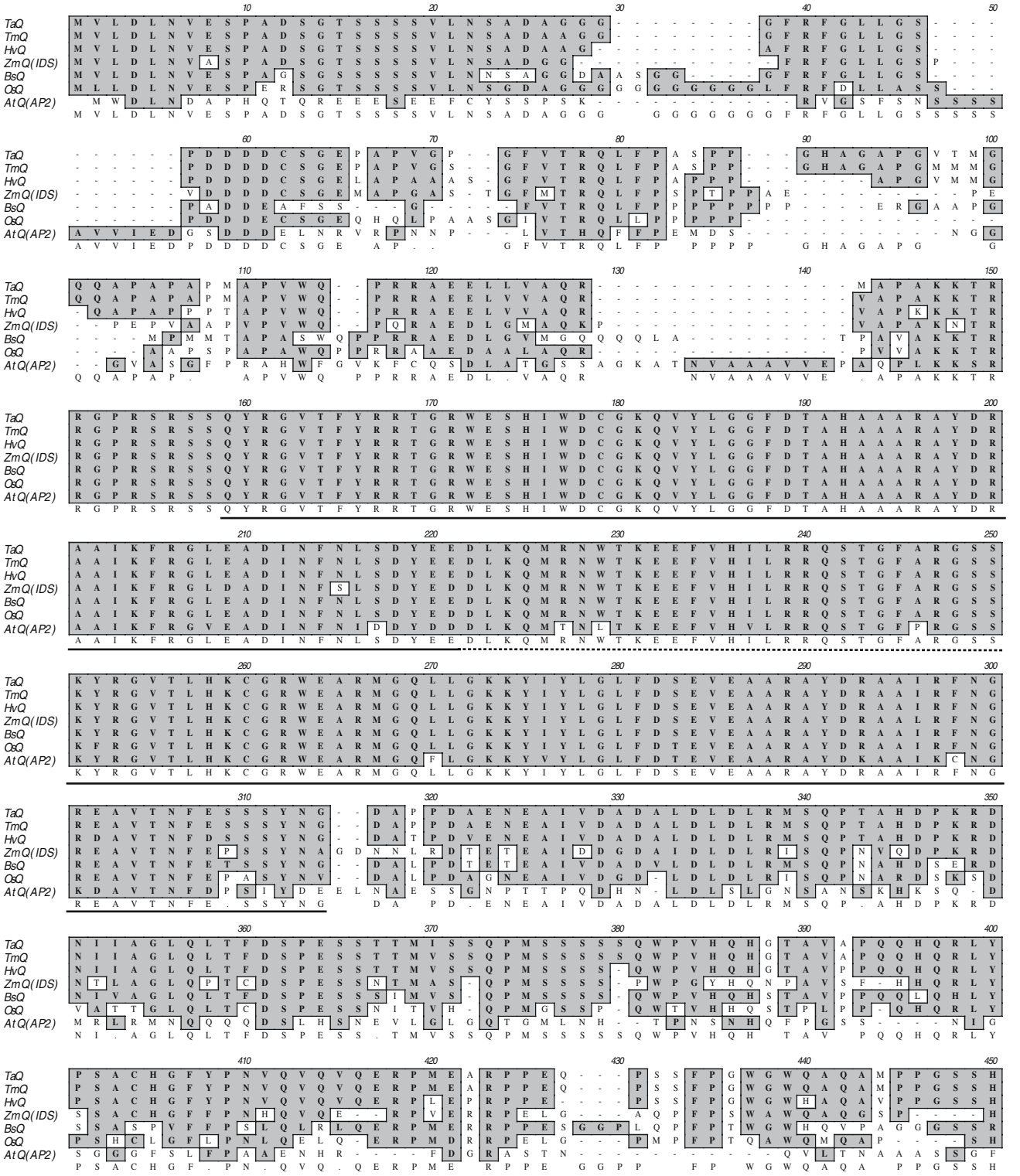
Amino acid sequences for comparative analysis of the *LH* genes were available from *Arabidopsis*, rice, *B. sylvaticum*, and *T. monococcum*. For the most part, the predicted amino acid sequences were highly conserved among the four species (Fig. 3). Amino acid identities ranged from 65% between *TmLH* and *AtLH* to 92% identity between *TmLH* and *BsLH*. The results of the phylogenetic analysis of the *LH* gene amino acid sequences agreed with those using the *Q* gene orthologs in that *Brachypodium* is more closely related to *T. monococcum* than is rice (Fig. 6).

Discussion

Brachypodium has been proposed as a model organism for the large-genome temperate grasses because of numerous desirable characteristics such as a small genome, rapid generation time, small size, amenability to transformation techniques, and its closer evolutionary relationship with the Triticeae than rice (Catalan and Olmstead 2000; Draper et al. 2001; Bennett and Leitch 2005; Vogel et al. 2006a, b; Bossolini et al. 2007). However, to date, few studies evaluating the level of microcolinearity between *Brachypodium*, rice, and the large-genome grasses such as wheat have been performed (Griffiths et al. 2006; Bossolini et al. 2007). Such studies are needed to provide knowledge regarding the potential utility of *Brachypodium* and its genomic sequence as a surrogate for map-based cloning and structural and functional genomics studies in wheat. The purpose of this study was to evaluate the level of microcolinearity between *Brachypodium*, rice, and wheat within the region harboring the *Q* locus. Although the functions of the corresponding *Q* gene orthologs in *Brachypodium* and rice are unknown, it is a major regulatory gene with pleiotropic effects on numerous development and domestication related traits in wheat (Leighty and Boshnakian 1921; Mackey 1954; Muramatsu 1963; 1986; Kato et al. 1999; Faris and Gill 2002; Faris et al. 2003; Simons et al. 2006). Therefore, this study provides analysis of microcolinearity and evolutionary events surrounding an essential locus of wheat.

Microcolinearity between *T. monococcum*, rice, and *B. sylvaticum*

In comparing microcolinearity between *T. monococcum* and rice, we found that all the genes present in the 50-kb rice segment were present in *T. monococcum* and were colinear. The only difference in microcolinearity of genes between *T. monococcum* and rice was the absence of an F-box-like ortholog from rice, which was present in four copies in *T. monococcum*. It is unknown if additional genes exist in *T. monococcum* between the *Tm40S* and *XBE406609* loci (orthologous to *OsMIIP*) that are not present in the rice segment because we do not have the complete physical


TaQ
TmQ
HvQ
ZmQ(IDS)
BsQ
OsQ
AtQ(AP2)

◀ **Fig. 5** Alignment of deduced amino acid sequences of the *Q* gene orthologs from *Triticum aestivum* (*TaQ*), *T. monococcum* (*TmQ*), barley (*HvQ*), maize (*ZmQ*(*ids*)), *Brachypodium sylvaticum* (*BsQ*), rice (*OsQ*), and *Arabidopsis* (*AtQ*(*AP2*)). *Black underlined regions* represent the two AP2 DNA binding domains, and the *dotted line* represents the linker between the two AP2 domains

sequence of this interval in *T. monococcum*. Only a few studies have shown comparable levels of microcolinearity between wheat and rice and examples include the *Vrn1* (Yan et al. 2003) and *Ha* (Chantret et al. 2004) loci.

There were a number of exceptions in the level of microcolinearity when comparing the *B. sylvaticum* segment with both rice and *T. monococcum*. First, the genes present in the *T. monococcum* contig are colinear with those in *B. sylvaticum* except that the *40S* gene is degenerated in *B. sylvaticum* and went undetected by ab initio gene finders. Only by probe hybridization and sequence alignments was a short region of homology to *Tm40S* identified in *B. sylvaticum*. Second, although *B. sylvaticum* contains an orthologous *FB* gene, there are four copies in *T. monococcum,* three of which are colinear. The most striking observation was that the *B. sylvaticum* segment harbors two additional genes, *BsKF* and *BsB1*, not found in the orthologous *T. monococcum* or rice segments. Genes with homology to *BsKF* and *BsB1* were found on rice chromosomes 11 and 2, respectively, and the wheat EST with homology to *BsKF* mapped to wheat homoeologous chromosome group 2. Although we cannot rule out the possibility that a wheat homolog of *BsKF* exists within the colinear region between *XBG263210* and *XBE406609*, it is evident that no homolog of the *BsB1* gene exists on the *T. monococcum* BAC contig between *TmQ* and *Tm40S* indicating that *BsB1* is not in a colinear position in *T. monococcum*. This would suggest that the translocation of these genes (at least *BsB1*) to this segment in *Brachypodium* occurred after the divergence of *Brachypodium* and the Triticeae.

This finding is contrary to that reported by Bossolini et al. (2007). In the evaluation of a 371-kb region of *B.*

Table 4 Percent amino acid identities between *Q* gene orthologs from hexaploid bread wheat (*TaQ*), diploid wheat (*TmQ*), barley (*HvQ*), *Brachypodium sylvaticum* (*BsQ*), maize (*ZmQ*), rice (*OsQ*), and *Arabidopsis* (*AtQ*)

Q ortholog	*TmQ*	*HvQ*	*BsQ*	*ZmQ(ids)*	*OsQ*	*AtQ(AP2)*
TaQ	98	89	75	72	71	42
TmQ		89	75	72	71	42
HvQ			74	72	71	43
BsQ				67	69	42
ZmQ(ids)					68	43
OsQ						42

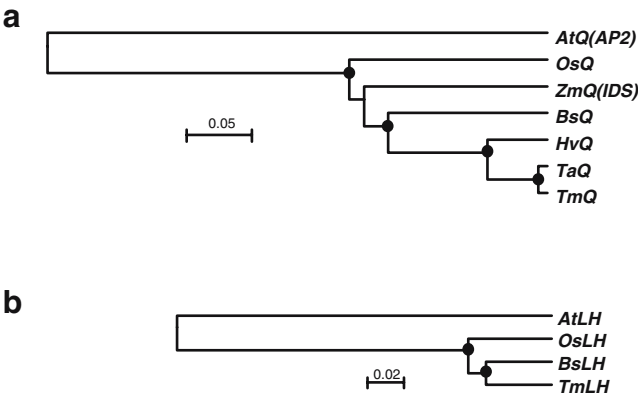


Fig. 6 Phylogenetic analysis of the predicted amino acid sequences of **a** the *Q* gene orthologs and **b** the leukotriene A-4 hydrolase (*LH*)-like gene orthologs. Trees were constructed from ClustalW alignments using the UPGMA method. Species are indicated with the prefixes *At* (*Arabidopsis thaliana*), *Os* (*Oryza sativa*), *Zm* (*Zea mays*), *Bs* (*Brachypodium sylvaticum*), *Hv* (*Hordeum vulgare*), *Tm* (*Triticum monococcum*), and *Ta* (*Triticum aestivum*). *Blackened circles* represent nodes supported by bootstrap values greater than 70%

sylvaticum and its comparison with wheat and rice, they found that five corresponding Triticeae BACs contained 15 genes, and only 11 of these were colinear with *B. sylvaticum*. Furthermore, among the 371-kb *B. sylvaticum* sequence, only one gene that was expected to be found among the Triticeae BACs was absent. Therefore, the study by Bossolini et al. (2007) suggests that when comparing wheat and *Brachypodium*, more noncolinear genes are expected to exist in wheat than *Brachypodium*. Our study indicates that *B. sylvaticum* has more noncolinear genes than both wheat and rice. It is possible that this contradiction occurred simply by chance and reflects the relatively small localized genomic regions under investigation. When the *Brachypodium* genome sequence becomes available in the near future, extensive quantitative analysis of microcolinearity and sequence conservation between *Brachypodium* and rice will be feasible and provide a better understanding of relationships at the whole-genome level.

Whereas *B. distachyon* is currently regarded as a true diploid, the exact genome composition of *B. sylvaticum* is less clear. Griffiths et al. (2006) reported finding two regions of the *B. sylvaticum* genome colinear with the *Ph1* region of wheat chromosome 5B, but one of the regions had a much better level of colinearity with the *Ph1* region than did the other. Therefore, it is possible that *B. sylvaticum* is an ancient or reduced tetraploid. This raises the possibility that more than one genomic region orthologous to the wheat *Q* gene region might exist in *B. sylvaticum*. However, screening of the *B. sylvaticum* BAC library with a fragment of the wheat *Q* gene as a probe revealed only six clones, which is the number expected for a single-copy gene when screening a library of 6.6 genome equivalents. Furthermore, all six clones had very similar fingerprints

indicating they were derived from a common genomic region. This suggests that the orthologous *Q* gene region is not duplicated in *B. sylvaticum*, and one would expect this to hold true for *B. distachyon* as well.

Gene density in *T. monococcum*, rice, and *B. sylvaticum*

Gene density is not static and known to vary dramatically across genomes. The orthologous *Q* regions in *B. sylvaticum* and rice were nearly identical but differed slightly in gene density. Gene density within the 50-kb rice segment under investigation was one gene per 10 kb, which agrees with the average for the whole genome (International Rice Genome Sequencing Project 2005). For the *B. sylvaticum* segment, average gene density was one gene per 14 kb, but all seven genes were clustered within the first 55 kb giving a higher density of one gene per 8 kb for that region. If we consider the average whole BAC gene density of one per 15 kb, our results agree with those of Bossolini et al. (2007) who reported a higher gene density for rice (one gene per 6 kb) compared to *B. sylvaticum* (one gene per 9 kb) in the orthologous *Lr34* region. The fact that all the *B. sylvaticum* genes were clustered within the first half of the BAC suggests that the *Brachypodium* genome consists of large tracts of high gene density, but interspersed among these tracts may be regions of repetitive elements.

No species of the Triticeae has been completely sequenced yet, so precise estimates of gene densities are not known. Average genome-wide gene densities are expected to be about one gene per 47 kb for *T. aestivum* (Stein 2007), but local sequencing has revealed densities as high as one gene per 5 kb (Feuillet and Keller 1999). Other regions of the Triticeae genomes are expected to be much lower in gene density. For example, Lu and Faris (2006) reported a 200-kb segment that contained no genes. Within the *Q* locus of *T. monococcum*, we observed a gene density of one gene per 41 kb, which is similar to the expected density over the whole genome. The density ranged from one gene per 14 kb to one per 133 kb, indicating that certain regions of the Triticeae genomes may have densities similar to that for *Brachypodium*. However, these estimates are based on consideration of the four copies of the *TmFB* gene, three of which are likely inactive. Taking into account that gene duplications tend to inflate density estimates and deflate intergenic distances, removal of the three likely inactive *TmFB* genes from the equation gives a more realistic functional gene density estimate of one gene per 73 kb with a range of one per 55 kb to one per 133 kb. The expanded intergenic distances observed in *T. monococcum* relative to *B. sylvaticum* and rice are due primarily to the presence of TEs, which are known to be inserted generally within the past few million years (SanMiguel et al. 2002).

Evolution of the *Q* locus

It is most likely that a common Poaceae ancestor harbored the six colinear genes *MIIP*, *PHD*, *40S*, *Q*, *FB*, and *LH*. Upon divergence of rice from the Triticeae/*Brachypodium* lineage, the *OsFB* gene was deleted from this locus with the resulting locus harboring the remaining five genes (Fig. 4). The Triticeae/*Brachypodium* lineage retained all six genes, but several events took place in the *Brachypodium* lineage after it diverged from the Triticeae, including the degeneration of the *40S* gene and likely transpositions/insertions of the *BsKF* and *BsB1* genes. In *T. monococcum*, the set of genes remained intact with the only apparent event being the duplication of *TmFB* multiple times. It is likely that these duplications exist among other members of the Triticeae as well because hybridization of the *TmFB1* gene to *T. aestivum* revealed that it detected multiple loci on each of the homoeologous group 5 chromosomes (data not shown).

Evaluation of the *Q* locus microcolinearity alone might suggest that rice is more closely related to *T. monococcum* than is *Brachypodium* because we observed more differences in microcolinearity between *T. monococcum* and *B. sylvaticum* than between *T. monococcum* and rice. However, a number of studies have demonstrated convincingly that *Brachypodium* is more closely related to the Triticeae than is rice (Catalan et al. 1995; Catalan and Olmstead 2000; Vogel et al. 2006b; Griffiths et al. 2006; Bossolini et al. 2007). Our comparative sequence and phylogenetic analysis of the *Q* and *LH* orthologs from the three species agree with this notion and show that *Brachypodium* is more closely related to the Triticeae than rice is. Together, these results demonstrate the importance of comparing more than two species to reveal true evolutionary relationships as indicated by Feuillet et al. (2001).

The potential of *Brachypodium* as a model for wheat genomics

The ideal model for wheat and other large-genome members of the Triticeae would contain all the Triticeae genes in a colinear order within a very small genome. The results of our study at the *Q* locus indicate that *Brachypodium* meets these criteria, except for the apparent degeneracy of the *40S* gene. This result agrees with previous studies that demonstrated high levels of conservation among genetic markers between *Brachypodium* and the Triticeae (Griffiths et al. 2006; Hasterok et al. 2006; Bossolini et al. 2007). However, in our study, *B. sylvaticum* also contained two genes in addition to all the *T. monococcum* genes, which suggests that like rice, numerous rearrangements and disruptions may have occurred

in *Brachypodium* relative to the Triticeae. These results together with those of Bossolini et al. (2007) suggest that *Brachypodium* would not be an ideal model for the large Triticeae genomes and that care and discretion would be needed if one were to use *Brachypodium* as a vehicle for map-based cloning in the Triticeae.

A much better understanding of the relationships between *Brachypodium* and the rest of the grass family will be obtained in the near future with the completion of the *B. distachyon* genome sequence. Although *Brachypodium* may not be suitable as a map-based cloning surrogate for wheat and barley, its sequence will provide a critical tool for comparative genomics, gene annotation, and the study of evolutionary relationships among the grasses. *Brachypodium* may also serve as a valuable model to study functional genomics of the grasses. Its amenability to transformation by *Agrobacterium*, rapid generation time, and small size make it an attractive candidate for such a purpose.

In conclusion, good levels of microcolinearity at the *Q* locus among *T. monococcum*, *B. sylvaticum*, and rice were observed, but more interruptions were observed between *T. monococcum* and *B. sylvaticum* than between *T. monococcum* and rice. Despite this, comparative sequence analysis of two colinear genes placed *Brachypodium* closer to *T. monococcum* on the evolutionary scale than rice, which agrees with other reports. Although more such studies are needed, this and the few other such studies that have been conducted suggest that the availability of the *Brachypodium* genome sequence will not preclude the need to conduct large-scale analysis of the Triticeae genomes. However, the availability of a complete genome sequence of rice and the soon-to-be-completed genome sequences of *B. distachyon* and sorghum will together provide Triticeae researchers a number of extremely valuable tools for genome analysis.

Acknowledgments This research was supported by USDA-ARS CRIS project 5442-22000-030.

References

Ahn S, Anderson JA, Sorrells ME, Tanksley SD (1993) Homoeologous relationships of rice, wheat and maize chromosomes. Mol Gen Genet 241:483–490

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bennett MD, Leitch IJ (2005) Nuclear DNA amounts in Angiosperms: progress, problems and prospects. Ann Bot 95:45–90

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

Bossolini E, Wicker T, Knobel PA, Keller B (2007) Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation. Plant J 49:704–717

Catalan P, Olmstead RG (2000) Phylogenetic reconstruction of the genus *Brachypodium* P. Beauv. (Poaceae) from combined sequences of chloroplast *ndh*F gene and nuclear ITS. Plant Syst Evol 220:1–19

Catalan P, Ying S, Armstrong L, Draper J, Stace CA (1995) Molecular phylogeny of the grass genus *Brachypodium* P. Beauv. based on RFLP and RAPD analysis. Bot J Linn Soc 117:263–280

Chantret N, Cenci A, Sabot F, Anderson O, Dubcovsky J (2004) Sequencing of the *Triticum monococcum hardness* locus reveals good microcolinearity with rice. Mol Gen Genomics 271:377–386

Chuck G, Meeley RB, Hake S (1998) The control of maize spikelet meristem fate by the *APETALA2*-like gene *indeterminate spikelet1*. Genes Dev 12:1145–1154

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

Distelfeld A, Uauy C, Olmos S, Schlatter AR, Dubcovsky J, Fahima T (2004) Microcolinearity between a 2-cM region encompassing the grain protein content locus *Gpc-6B1* on wheat chromosome 6B and a 350-kb region on rice chromosome 2. Funct Integr Genomics 4:59–66

Draper J, Mur LAJ, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, Routledge APM (2001) *Brachypodium dystachyon*: a new model system for functional genomics in grasses. Plant Phys 127:1539–1555

Faris JD, Gill BS (2002) Genomic targeting and high-resolution mapping of the domestication gene *Q* in wheat. Genome 45:706–718

Faris JD, Haen KM, Gill BS (2000) Saturation mapping of a gene-rich recombination hot spot in wheat. Genetics 154:823–835

Faris JD, Fellers JP, Brooks SA, Gill BS (2003) A bacterial artificial chromosome contig spanning the major domestication gene *Q* in wheat and identification of a candidate gene. Genetics 164:311–321

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, Keller B (2002) Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. Ann Bot 89:3–10

Feuillet C, Penger A, Gellner K, Mast A, Keller B (2001) Molecular evolution of receptor-like kinase genes in hexaploid wheat. Independent evolution of orthologs after polyploidization and mechanisms of local rearrangements at paralogous loci. Plant Physiol 125:1304–1313

Foote TN, Griffiths S, Allouis S, Moore G (2004) Construction and analysis of a BAC library in the grass *Brachypodium sylvaticum*: its use as a tool to bridge the gap between rice and wheat in elucidating gene content. Funct Integr Genomics 4:26–33

Francki M, Carter M, Byan K, Hunter A, Bellgard M, Appels R (2004) Comparative organization of wheat homoeologous group 3S and 7L using wheat-rice synteny and identification of potential markers for genes controlling xanthophyll content in wheat. Funct Integr Genomics 4:118–130

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

Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore M (2006) Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. Nature 439:749–752

Hasterok R, Marasek A, Donnison IS, Armstead I, Thomas A, King IP, Wolny E, Idziak D, Draper J, Jenkins G (2006) Alignment of the genomes of *Brachypodium distachyon* and temperate cereals

and grasses using bacterial artificial chromosome landing with fluorescence *in situ* hybridization. Genetics 173:349–362

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Kato K, Miura H, Sawada S (1999) QTL mapping of genes controlling ear emergence time and plant height on chromosome 5A of wheat. Theor Appl Genet 98:472–477

Lagudah ES, McFadden H, Singh RP, Huerta-Espino J, Bariana HS, Spielmeyer W (2006) Molecular genetic characterization of the *Lr34/Yr18* slow rusting resistance gene region in wheat. Theor Appl Genet 114:21–30

Lander ES, Green P, Abranhamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Leighty CE, Boshnakian S (1921) Genetic behaviour of the spelt form in crosses between *Triticum spelta* and *Triticum aestivum*. J Agric Res 7:335–364

Li W, Gill BS (2002) The colinearity of the *Sh2/A1* orthologous region in rice, sorghum and maize is interrupted and accompanied by genome expansion in the Triticeae. Genetics 160:1153–1162

Lu HJ, Faris JD (2006) Macro- and micro-colinearity between the genomic region of wheat chromosome 5B containing the *Tsn1* gene and the rice genome. Funct Integr Genomics 6:90–103

MacKey J (1954) Neutron and X-ray experiments in wheat and a revision of the speltoid problem. Hereditas 40:65–180

Muramatsu M (1963) Dosage effect of the spelta gene q of hexaploid wheat. Genetics 48:469–482

Muramatsu M (1986) The *vulgare* super gene, *Q*: its universality in *durum* wheat and its phenotypic effects in tetraploid and hexaploid wheats. Can J Genet Cytol 28:30–41

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals and consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903–9908

SanMiguel PJ, Ramakrishna W, Bennetzen JL, Busso C, Dubcovsky J (2002) Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome $5A^m$. Funct Integr Genomics 2:70–80

Schnurbusch T, Collins NC, Eastwood RF, Sutton T, Jefferies SP, Langridge P (2007) Fine mapping and targeted SNP survey using rice-wheat gene colinearity in the region of the *Bo1* boron toxicity tolerance locus of bread wheat. Theor Appl Genet 115:451–461

Sears ER (1954) The aneuploids of common wheat. MO Agr Exp Sta Res Bull 572:1–59

Simons KJ, Fellers JP, Trick HN, Zhang Z, Tai Y-S, Gill BS, Faris JD (2006) Molecular characterization of the major wheat domestication gene *Q*. Genetics 172:547–555

Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin, Mahmoud A, Ma X, Gustafson PJ, Qi LL, Echalier B, Gill BS, Matthews DE, Lazo GR, Chao S, Anderson OD, Edwards H, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorak J, Zhang D, Nguyen HT, Peng J, Lapitan NLV, Gonzalez-Hernandez JL, Anderson JA, Hossain K, Kalavacharla V, Kianian SK, Choi DW, Close TJ, Dilbirligi M, Gill KS, Steber C, Walker-Simmons MK, McGuire PE, Qualset CO (2003) Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 13:1818–1827

Stein N (2007) Triticeae genomics: advances in sequence analysis of large genome cereal crops. Chromosome Res 15:21–31

Tatusova TA, Madden TL (1999) BLAST 2 Sequences: a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett 174:247–250

Valarik M, Linkiewicz AM, Dubcovsky J (2006) A microcolinearity study at the earliness *per se* gene *Eps-A^m1* region reveals an ancient duplication that preceded the wheat-rice divergence. Theor Appl Genet 112:945–957

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

Van Deynze AE, Nelson JC, Yglesis ES, Harrington SE, Braga DP, McCouch SR, Sorrells ME (1995b) Comparative mapping in grasses. Wheat relationships. Mol Gen Genet 248:744–754

Vogel JP, Garvin DF, Leong OM, Hayden DM (2006a) *Agrobacterium*-mediated transformation and inbred line development in the model grass *Brachypodium distachyon*. Plant Cell Tissue Organ Cult 85:199–211

Vogel JP, Gu Y-Q, Twigg P, Lazo GR, Laudencia-Chingcuanco D, Hayden DM, Donze TJ, Vivian LA, Stamova B, Coleman-Derr D (2006b) EST sequencing and phylogenetic analysis of the model grass *Brachypodium distachyon*. Theor Appl Genet 113:186–195

Ware DH, Jaiswal P, Ni J, Yap IV, Pan X, Clark KY, Teytelman L, Schmidt SC, Zhao W, Chang K, Cartinhour S, Stein LD, McCouch SR (2002) Gramene, a tool for grass genomics. Plant Physiol 130:1606–1613

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of wheat vernalization gene *VRN1*. Proc Natl Acad Sci USA 100:6263–6268