Title

Sequencing of the Triticum monococcum Hardness locus reveals good microcolinearity with rice

Journal

Molecular Genetics and Genomics, 271(4)

ISSN

1617-4615

Authors

Chantret, N
Cenci, A
Sabot, F
et al.

Publication Date

2004-05-01

DOI

10.1007/s00438-004-0991-y

Peer reviewed

Mol Gen Genomics (2004) 271: 377–386
DOI 10.1007/s00438-004-0991-y

ORIGINAL PAPER

N. Chantret · A. Cenci · F. Sabot · O. Anderson
J. Dubcovsky

Sequencing of the *Triticum monococcum Hardness* locus reveals good microcolinearity with rice

Received: 17 December 2003 / Accepted: 16 February 2004 / Published online: 11 March 2004

Abstract The *Hardness* (*Ha*) locus on chromosome 5D is the main determinant of grain texture in hexaploid wheat. The related genes *Puroindoline-a* and *-b* (*Pina-D1*, *Pinb-D1*) and *Grain Softness Protein* (*Gsp-D1*) are tightly linked at this locus. Mutations in the *Pina-D1* and *Pinb-D1* genes are associated with increased grain hardness. We report here the complete sequence of a 101-kb BAC clone from *Triticum monococcum* (A^m genome) which includes these three genes, and its comparison with the orthologous region in rice. The genes *Gsp-A ^{m}1*, *Pina-A ^{m}1* and *Pinb-A ^{m}1* are separated by 37 kb and 32 kb, respectively, and are organized in the same transcriptional orientation. Four additional genes, including a pair of duplicated genes, were identified upstream of *Gsp-A ^{m}1* within a high-density gene island. These additional genes were found in the same order and orientation, and the same relative distances apart as similar genes previously annotated on rice chromosome 12. An interesting discovery was a small unannotated putative rice gene that was similar to the *Gsp-A ^{m}1* gene of *T. monococcum* (65% similarity at the protein level), and that was disposed in the same orientation, and located in the same position relative to the other orthologous genes. The high gene density observed in this BAC (1 gene per 14 kb) was expected for a distal chromosome region, but the level of microcolinearity with rice was higher than that reported in similar distal regions of other wheat chromosomes. Most of the BAC sequence (40%) was represented by repetitive elements, mainly concentrated in regions adjacent to the genes *Pina-A ^{m} 1* and *Pinb-A ^{m} 1*. Rearrangements among these repetitive elements might provide an explanation for the frequent deletions observed at this locus in the genomes of the polyploid wheat species.

Keywords Wheat · *Hardness* · Grain Softness Protein (GSP) · Puroindoline · Microcolinearity

Communicated by A. Kondorosi

N. Chantret
INRA, CIRAD-AMIS-Biotrop, Av. Agropolis, TA 40/30, 34398 Cedex 5 Montpellier, France

A. Cenci
Dipartimento di Biologia e Chimica Agroforestale e Ambientale, Università degli Studi di Bari, Via G. Amendola 165/a, 70126 Bari, Italy

F. Sabot
INRA, Station d'Amélioration des Plantes, Domaine de Crouelle, 63000 Clermont-Ferrand, France

O. Anderson
Western Region Research Center, USDA, 800 Buchanan St., Albany, CA 94710, USA

J. Dubcovsky (✉)
Dept. of Agronomy and Range Science, University of California, Davis, CA 95616, USA
E-mail: jdubcovsky@ucdavis.edu
Tel.: +1-530-7525159
Fax: +1-530-7524361

Introduction

The grass family (*Poaceae*) includes many economically important species such as rice, maize, sorghum, oat, barley, rye and wheat. This family originated approximately 55–70 million years ago (Kellogg 2001) and includes a very diverse set of species. This diversity is manifested in huge differences in nuclear DNA content, which varies from 430 Mb in rice to 5700 Mb in diploid wheat *Triticum monococcum* (Arumuganathan and Earle 1991). Despite these large differences in DNA content, relatively good gene colinearity was observed in the first comparative RFLP maps constructed for wheat and rice (Moore et al. 1995; Van Deynze et al. 1995).

More recently, orthologous regions from the small rice genome and the large genomes of wheat and barley have been compared at the sequence level (microcolinearity) (Panstruga et al. 1998; Shirasu et al. 2000; Dubcovsky et al. 2001; Wicker et al. 2001; Brooks et al. 2002; Li and Gill 2002; Ramakrishna et al. 2002; Brunner et al. 2003; Yan et al. 2003). A feature common

to all these comparisons was the lack of sequence conservation in the intergenic regions. This rapid divergence of intergenic regions was also observed in comparisons between different genera within the Triticeae (Ramakrishna et al. 2002; Gu et al. 2003), and even in comparisons between different *Triticum* species that diverged only 1–3 million years ago (Myr) (Wicker et al. 2003b). Significant differences in intergenic regions have been recently reported in different maize genotypes of the same species (Fu and Dooner 2002).

In general, intergenic regions are larger in the Triticeae genomes than in rice, due to the presence of repetitive elements, which are usually found to form nested structures (Panstruga et al. 1998; Dubcovsky et al. 2001; Wicker et al. 2001; Brooks et al. 2002; Rostoks et al. 2002; SanMiguel et al. 2002). Retrotransposons represent the most abundant class of repetitive elements, but other families, such as CACTA transposons, have also contributed significantly to the large size of the wheat and barley genomes (Wicker et al. 2003a). At a smaller scale, local duplications and insertion of non-retroelement sequences have also contributed to genome size increase (Feuillet et al. 2001; Wicker et al. 2001). Increases in genome size are counterbalanced by the elimination of large DNA segments by unequal recombination between LTRs of the same retrotransposon generating solo-LTRs, unequal recombination between paralogous related retroelements, and other mechanisms (Shirasu et al. 2000; Devos et al. 2002).

In contrast to the complete divergence of the intergenic regions, relatively good genic microcolinearity has been observed in most of the comparisons between rice and the Triticeae species. However, microcolinearity is usually interrupted by gene deletions, translocations, inversions and/or duplications (for reviews, see Bennetzen 2000; Keller and Feuillet 2000; Bennetzen and Ramakrishna 2002; Feuillet and Keller 2002). Exceptions to microcolinearity have been observed also among the more closely related wheat genomes, although at lower frequencies. The *Hardness* (*Ha*) locus is an example of an agronomically important chromosomal region that has suffered numerous independent deletions during wheat evolution (Giroux and Morris 1998; Gautier et al. 2000; Tranquilli et al. 2002). The relatively recent occurrence of these deletions, and their economic importance, make them an interesting case in which to study these phenomena.

Wheat grain hardness, defined as either the resistance of kernels to crushing or the particle size distribution in the flour, is an important trait that determines the end-use quality of wheat (reviewed in Anjum and Walker 1991; Morris 2002). In bread wheat, endosperm texture is primarily controlled by the *Hardness* (*Ha*) locus on the short arm of chromosome 5D (Mattern et al. 1973; Law et al. 1978). It is a simply inherited character (Symes 1965) and although the main locus is referred as "*Hardness*", softness is in fact the dominant trait. Soft wheat varieties carrying the dominant *Ha* allele have a higher abundance of a starch surface-associated protein ($M_r = 15$ kDa) named friabilin (Greenwell and Schofield 1986; Rahman et al. 1994). Friabilin is a composite of related lipid-binding proteins including proteins coded by the genes *Puroindoline-a* (*Pina*) and *Puroindoline-b* (*Pinb*) and, in lower proportion, the protein coded by the *Grain Softness Related* (*Gsp*) gene (Gautier et al. 1994; Rahman et al. 1994). The puroindoline proteins are more similar to each other (71%) than to the Grain Softness Related Protein (Gsp, 57–58%). These three related genes are physically close to each other in both the A^m and D genomes (Tranquilli et al. 1999; Turnbull et al. 2003), suggesting that they originated by two independent duplication events (Tranquilli et al. 1999). Studies of natural mutations (Morris 2002) and deletions (Tranquilli et al. 2002), and work on transgenic wheat (Beecher et al. 2002), have confirmed that the puroindoline genes are the ones responsible for the differences in grain texture. Altered dosages of the *Gsp* genes do not affect grain hardness (Tranquilli et al. 2002).

Puroindoline genes have been found in the diploid ancestors of wheat, as well as in barley (hordoindoline), rye (secaloindoline), and oat (avenoindoline), but not in tetraploid wheat or in the A and B genomes of bread wheat (Gautier et al. 2000). In addition, many common wheat varieties (*T. aestivum* L.) with hard-textured grains have deletions for the *Pina-D1* gene (Giroux and Morris 1998), whereas a few show a complete deletion of the three linked genes on chromosome 5D (Tranquilli et al. 2002). These results indicate that these genes are not essential for the normal growth of cultivated wheat—an idea that is also supported by the previous failure to detect orthologues of these genes in rice, maize or sorghum (Gautier et al. 2000). Transgenic rice expressing puroindoline-a and/or -b showed significantly increased tolerance to *Magnaporthe grisea* (rice blast), displaying a 29–54% reduction in symptoms, and to *Rhizoctonia solani* (sheath blight), with an 11 to 22% reduction in symptoms (Krishnamurthy et al. 2001).

In the present paper we report an analysis of the complete sequence of a *T. monococcum* BAC clone that includes the *Pina-*A^m*1*, *Pinb-*A^m*1*, and *Gsp-*A^m*1* genes (Tranquilli et al. 1999), which provides a template for future comparison of the different deletions that have occurred in this locus during wheat evolution. In addition, we present a detailed comparison with rice, and report the discovery of a putative orthologue of the *Gsp* gene in rice.

Materials and methods

BAC selection and sequencing

The BAC clone 109N23 was isolated from a *T. monococcum* BAC library (Lijavetzky et al. 1999), and includes the genes *Pina-*A^m*1*, *Pinb-*A^m*1*, and *Gsp-*A^m*1* (Tranquilli et al. 1999). The clone was sequenced as described before (Dubcovsky et al. 2001). A small difference from the previous protocol was the reduced size of the fragments in the shotgun library (2.5 kb, on average). Base calling

and sequence quality assessments were carried out using PHRED (Ewing and Green 1998) and sequence assembly was done with PHRAP and edited with CONSED (Gordon et al. 1998). Gaps were filled first by re-sequencing subclones indicated by the 'autofinish' option of CONSED and by primer walking of subclones bridging the sequencing gaps. The final error rate was estimated using CONSED.

Sequence analysis

Detection of ORFs and determination of gene structures were performed using a combination of gene-finding programs and BLAST analyses. The gene-finding programs used were Genemark.hmm with rice and wheat training sets (Lukashin and Borodovsky 1998; http://opal.biology.gatech.edu/GeneMark/eukhmm.cgi) and GeneScan with the maize training set (Burge and Karlin 1997; http://genes.mit.edu/GENSCAN.html). Sequences were also compared to NCBI non-redundant databases and dbEST, and to the TIGR whole rice genome database (http://tigrblast.tigr.org/euk-blast/index.cgi?project=osa1) using BLASTN, BLASTP, BLASTX and TBLASTX algorithms (Altschul et al. 1997). When both predictors and BLAST analysis gave coherent indications regarding the presence of a gene, the final intron/exon structure was predicted from the alignment of the genomic sequence with Triticeae ESTs (when available) or with the putatively orthologous rice gene. The automatic rice BAC annotation generated by TIGR was rechecked by aligning the rice genomic sequence with rice ESTs and/or by comparison with the wheat BAC sequence. The NetPlantGene splice-site prediction program was used to determine the level of coding potential for some particular regions (Hebsgaard et al. 1996; http://www.cbs.dtu.dk/services/NetPGene/).

Repeated sequences were identified with BLASTN, BLASTX and TBLASTX in the NCBI non-redundant (nr) database, TIGR (http://tigrblast.tigr.org/euk-blast/index.cgi?project=osa1), the Triticeae Repeat Sequence Database from GrainGenes (TREP; http://wheat.pw.usda.gov/ITMI/Repeats/index.shtml), and RepBase (Jurka 2000; http://www.girinst.org), as well as with the LTR_STRUC program (McCarthy and McDonald 2003). The Blast2Sequence program (Tatusova and Madden 1999) was used to compare the complete BAC sequence with itself.

Calculation of retrotransposon insertion times

The calculation of retrotransposon insertion times was done as described by SanMiguel et al (2002). The 5′ and 3′ LTRs of the four complete retrotransposons present in BAC 109N23 (Angela_109N23-2, Angela_109N23-3, Sabrina_109N23-1 and Retrofit_109N23-1) were aligned using ClustalX (Thompson et al. 1997). The Kimura two-parameter distances were calculated after having removed all sites containing alignment gaps (Option: 'Complete-deletion') as implemented in MEGA version 2.1 (Kumar et al. 2001). The insertion times was dated using as the mutation rate the estimate of 6.5×10^{-9} substitutions per synonymous site per year obtained by Gaut et al (1996) based on the *Adh1* and *Adh2* loci of grasses.

Results

BAC sequencing

The *T. monococcum* BAC clone 109N23 was sequenced at a target coverage level of 12X (1152 subclones were sequenced in both directions). The assembly resulted in 10 gaps and several stretches of poor-quality sequence. Approximately 96 subclones were re-sequenced as suggested by the 'autofinish' option of CONSED, closing three gaps. The seven remaining gaps and 15 regions with low-quality sequences were re-sequenced using primer walking. The final contiguous sequence of 101,101 bp (44.98% G + C) was deposited in GenBank under the Accession No. AY491681.

The assembled sequence of BAC 109N23 was validated experimentally by comparing it to a restriction map constructed with restriction enzymes with 8-bp recognition sites (*Not*I, *Pac* I, *Pme* I and *Asc* I). The restriction map obtained by Tranquilli et al. (1999) was almost identical to the sequence prediction. The only exceptions were the relative order of the four successive *Asc* I sites that could not be determined unequivocally in the experimental map, and the presence of two additional *Not* I sites located 55-bp and 158-bp from each other, yielding restriction fragments that would have been too small to be detected in the pulsed field electrophoresis gels used to construct the experimental restriction map.

Gene annotation

The three related genes *Gsp-A^{m}1* (25,946–26,440 bp), *Pina-A^{m}1* (63,533–63,979 bp), *Pinb-A^{m}1* (96,193–96,639 bp) are organized in the same orientation within a 70-kb region. *Gsp-A^{m}1* and *Pina-A^{m}1* are 37 kb apart, and *Pina-A^{m}1* and *Pinb-A^{m}1* are 32 kb apart, and the genes are separated mainly by repetitive elements (Fig. 1). Four additional genes were identified upstream of the first of these three genes, and were designated 1a (11,773–12,420-bp), 1b (17,293–17,961-bp), 2 (18,452–19,921-bp) and 3 (20,896–24,110-bp) to facilitate further discussion (Table 1 and Fig. 1). These genes were validated by alignment with ESTs and previously annotated genes, and by their conservation in the orthologous region of the rice genome.

Genes 1a and 1b (648 and 669 bp long, respectively) have no introns, and are located 5 kb apart, forming a duplication that includes 486 bp and 464 bp of sequence upstream and downstream of genes 1a and 1b, respectively (see Fig. 1). The percentage identity between the members of the complete duplication is 81.3% at the nucleotide level and 79.1% at the amino acid level for the predicted proteins. A cDNA from rye (BQ160244) was found to be 94% identical to gene 1b, confirming that it is an expressed sequence ($E=e^{-105}$). Conservation of this region with a hypothetical gene on rice chromosome 12 (BAC OSJNBb0011N16, 3,252–3,921-bp) supported the annotation of these genes. The degree of identity between the rice and wheat genes 1a and 1b was, respectively, 45 and 42% at the amino acid level. However, the predicted rice gene included one intron. The absence of an intron in the wheat genes was confirmed by the alignment with rye cDNA BQ160244. The similarity with the rye EST confirmed that this is an expressed sequence, although it probably not highly expressed since none of the 914,000 wheat or barley

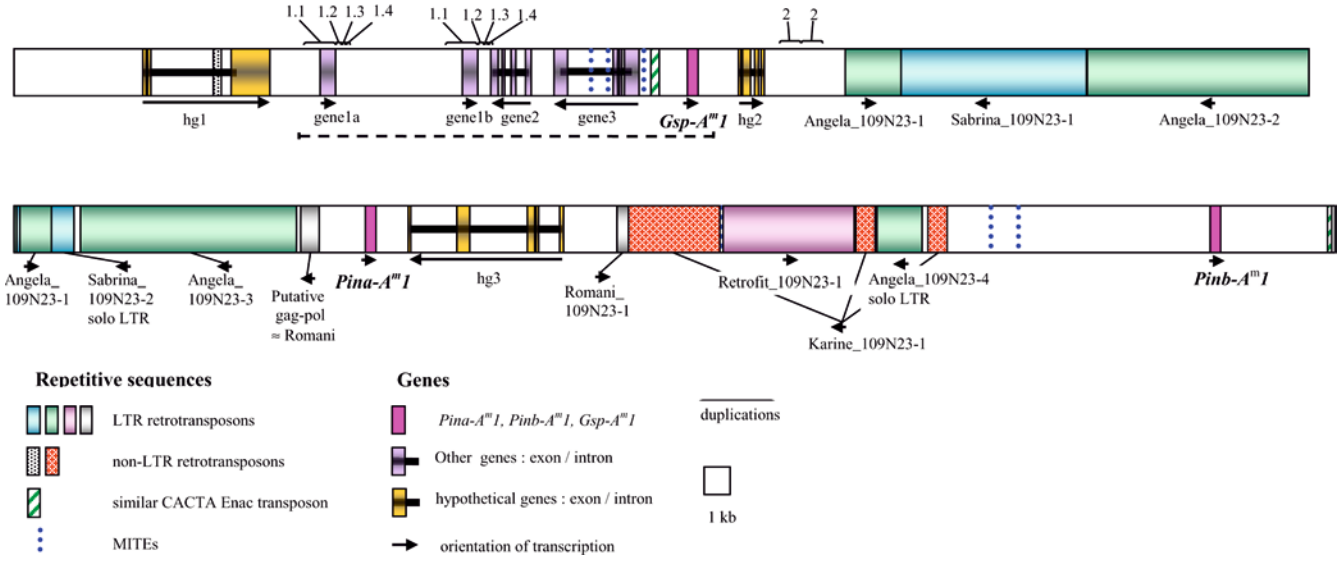

Fig. 1 Schematic representation of the annotation of the sequence of BAC 109N23 from *T. monococcum* (GenBank Accession No. AY491681). Genes are represented by *purple boxes* and hypothetical genes (hg) by *yellow boxes*. The *arrows* indicate the transcriptional orientation. LTR retrotransposons elements are indicated by *blue*, *red*, *gray* or *pink boxes*; full element names are given according to the nomenclature in the TREP database (Name_BAC number-rank in the sequence); non-LTR retrotransposons are represented by *gray* and *red dotted boxes*; transposons are represented *by green striped white boxes*; *blue interrupted bars* represent MITEs, *horizontal lines* represent direct repeats identified by a number. The *broken horizontal line* represents the part of the BAC detailed in Fig. 2

ESTs available at the time of this analysis showed similar levels of identity to the rye cDNA.

Gene 2 is 1470 bp long and has four exons. Several wheat cDNAs were significantly similar ($5e^{-91} \leq E \leq 2e^{-53}$) to this gene, confirming the annotation of the third and fourth exons. The complete intron-exon structure of the gene was validated by its similarity with its orthologous gene on rice chromosome 12 (BAC OSJNBb0011N16, 4,232–6,516-bp). The annotated rice gene structure was corrected to maximize the similarity with the wheat protein. A comparison of the predicted wheat protein with the SwissProt database (BLASTP)

Table 1 Characteristics of the predicted genes in the *T. monococcum* BAC 109N23 (GenBank AY491681)

Name	Strand	Position	Gene length (bp)	Exons	Protein size (aa)	Description	Best homology found and/or used (E value)
Gene1a	+	11773–12420	648	1	215	Putative protein	Rice BAC AL928743 gene 1, 'hypothetical protein'
Gene1b	+	17293–17961	669	1	222	Putative protein	Rice BAC AL928743 gene 1, 'hypothetical protein'/ Rye cDNA BQ160244 (e^{-105})
Gene2	–	18452–19921	1470	4	218	Synaptobrevin	Rice BAC AL928743 gene 2, 'putative synaptobrevin'/ Wheat cDNA CA670055 ($5e^{-91}$)
Gene3	–	20896–24110	3215	4	425	N-acetylglucosamine transferase	Rice BAC AL928743 gene 3, 'glycosylation enzyme-like protein'/ Wheat ESTs BU100707 (0.0), BE591544 ($7e^{-83}$), BE400153 ($3e^{-82}$)
Gsp- A^m1	+	25946–26440	495	1	164	Grain Softness Protein	-
Pina-A^m1	+	63533–63979	447	1	148	Puroindoline-a	-
Pinb-A^m1	+	96193–96639	447	1	148	Puroindoline-b	-
Hypothetical gene1 (hg1)	+	4919–9900	4982	3	553	Hypothetical gene	No clear orthologues in other plants and no similar ESTs, but low similarity to Rice Protein BAB64674 ($2e^{-09}$)
Hypothetical gene2 (hg2)	+	27995–28967	973	4	240	Hypothetical gene	No clear orthologues in other plants, but partial similarity to wheat EST BG606188 ($5e^{-34}$) [a]
Hypothetical gene3 (hg3)	–	65446–71210	1138	5	345	Hypothetical gene	No clear orthologues in other plants, but partial similarity to wheat EST CA498167 (0.0) [a]

[a]Similarity between a hypothetical gene and a Triticeae EST does not constitute sufficient proof of a functional gene, because numerous transposons are present in the wheat and barley EST collections (Echenique et al. 2002).

revealed significant similarity ($E = e^{-73}$) with a synaptobrevin-related protein from *Arabidopsis thaliana*; synaptobrevins are vesicle-associated membrane proteins.

Gene 3 is 3215 bp long and has four predicted exons that were confirmed by their almost perfect identity (≥95%, $3e^{-82} \leq E \leq 0.0$) with several wheat ESTs (BU100707, BE591544, BE400153) covering the complete predicted gene and partially covering both UTRs. Gene 3 also showed good similarity to its orthologous gene on rice chromosome 12 (BAC OSJNBb0011N16, 7,464–9,361). The rice gene had the same intron-exon structure, and the overall alignment of the wheat and rice protein showed 81% identity. A comparison of the predicted wheat protein with the Swissprot database (BLASTP) revealed significant similarity ($E = e^{-149}$) with *Arabidopsis thaliana* proteins belonging to the N-acetylglucosaminyltransferase family (Core-2/I-Branching enzyme family).

Gene prediction programs suggested the presence of three other hypothetical genes (hg), designated hg1 (4,919–9,900-bp), hg2 (27,995–28,967-bp) and hg3 (65,446–71,210-bp, Table 1; Fig. 1), which could not be unequivocally confirmed as functional genes by other methods (Table 1). Their predicted products showed either no similarity with any protein present in the non-redundant GenBank database (hg2), or low similarity with one (hg1) or multiple (hg3) predicted proteins from rice encoded on different rice chromosomes (putative uncharacterized repetitive elements).

If these three hypothetical genes are added to the seven validated genes discovered in BAC 109N23, the resulting gene density would be 1 gene per 10 kb. However, a more conservative calculation which excludes the three hypothetical genes results in an estimated gene density of 1 gene per 14 kb (7 genes in 101 kb).

Intergenic regions

Different kinds of repeat sequences were found in BAC 109N23, which account for 42% of the complete clone. The most abundant sequences were LTR retrotransposons, which make up 36% of the BAC sequence. Three complete LTR retrotransposons were found between the genes *Gsp-A^{m}1* and *Pina-A ^{m}1*, and one between *Pina-A ^{m} 1* and *Pinb-A^{m}1* (Fig. 1). The first one was the *Gypsy* -like element Sabrina_109N23-1 inserted in the partial *Copia* -like element Angela_109N23-1. The second one was another Angela_109N23-2, which was inserted into the 5′LTR of Sabrina_109N23-1. The third one, another complete Angels element (Angela_109N23-3), was flanked by a Sabrina solo-LTR (Sabrina_109N23-2) and a repetitive element with low similarity to a *Gypsy* -like *Romani* retrotransposon. The last one, Retrofit_109N23-1, is a *Copia* -like element, which is highly similar to the rice element *Retrofit* (E = 0.0 with TBLASTX) and has a relatively short LTR (196 bp). Insertion times calculated for the four complete LTR-retrotransposons showed that Retrofit_109N23-1 (0.4 ± 0.39 Myr) was the most recent insertion (only one SNP is observed between the 196-bp LTRs) and Sabrina_109N23-1 (3.1 ± 0.4 Myr) was the oldest (Table 2).

The region between the *Pina-A ^{m} 1* and *Pinb-A ^{m}1* genes also included numerous repetitive elements. A partial Romani (Romani_108N23-1) was followed by the non-LTR retrotransposon Karine (Karine_109N23-1), which has insertions of the Retrofit_109N23-1 and an Angela_109N23-4 solo-LTR. Two regions showed similarity to CACTA Enac transposons ($E = 5e^{-35}$), one after gene 2 and the other at the end of the BAC. Five MITEs (Miniature Inverted-repeat Transposable Elements) from the *Stowaway* family and one additional Fold back element were also found within this BAC.

Besides the annotated retroelements, additional direct duplications were found after genes 1a and 1b, and after the hypothetical gene hg2. A similar set of three short direct duplications (designated 1.2, 1.3 and 1.4 in Fig. 1) was found after genes 1a and 1b, suggesting that the duplication event included approximately 1.6 to 1.8 kb. The number 2 in Fig. 1 indicates the location of an additional direct repeat after the hypothetical gene hg2.

Assuming that regions of 1 kb upstream, and 0.5 kb downstream of each gene contain putative UTRs and regulatory regions, only 20% of the sequence (20 kb) showed no homology to any known gene or repetitive element and remained uncharacterized.

Colinearity between wheat and rice sequences

Wheat genes 1a, 1b, 2 and 3 have orthologues in rice, which are arranged in the same order and orientation,

Table 2 Description and time of insertion (in Myr) of four retrotransposons found in the *T. monococcum* BAC 109N23 (GenBank AY491681)

Designation	Length (bp)	Strand	Size of 5'LTR	Size(s) of deletion(s) in 5'LTR	Size of 3'LTR	Size(s) of deletion(s) in 3'LTR	Estimated age of insertion (Myr)	S.D.[a]
Angela_109N23-2	8726	–	1732	1	1728	3; 2	1.49	0.26
Angela_109N23-3	8391	–	1677	45; 45; 1	1766	1; 1	2.06	0.31
Sabrina_109N23-1	7261	–	1578	1; 1; 5	1571	1; 1; 11; 1	3.10	0.40
Retrofit_109N23-1	5194	+	196	None	196	None	0.40	0.39

[a]Standard deviation of age estimate

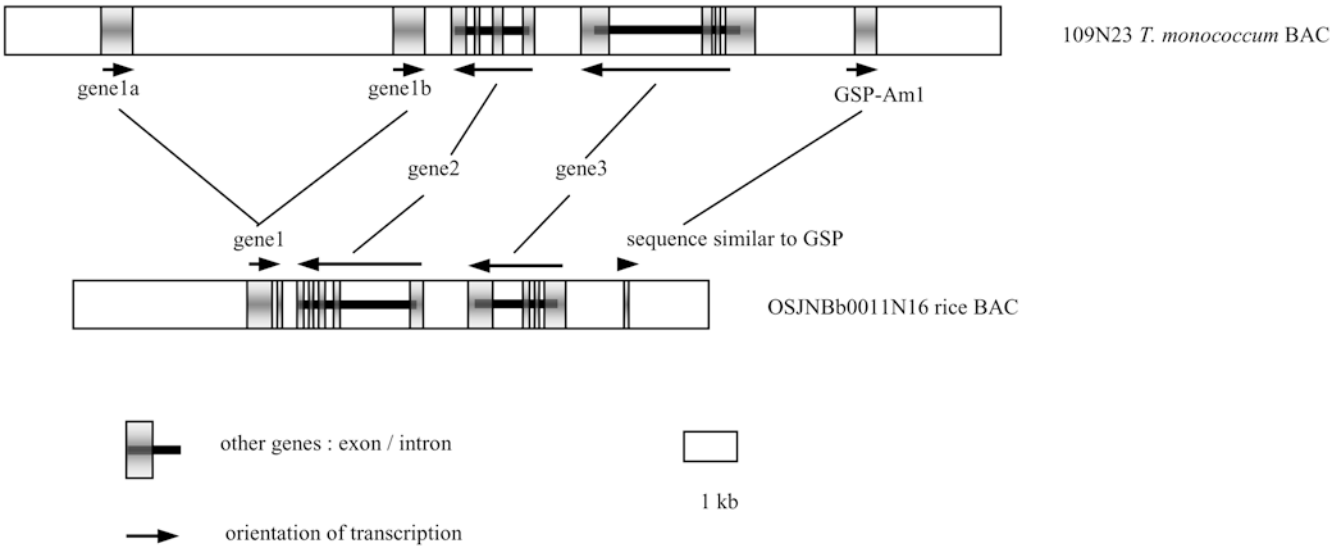

Fig. 2 Schematic representation of the microcolinearity between *T. monococcum* BAC 109N23 (GenBank AY491681) and rice BAC OSJNBb0011N16. Genes are represented by *gray boxes*; the *arrow* indicates the orientation of transcription

although gene 1 is duplicated in wheat but not in rice (Fig. 2). The distances among these genes are also conserved in both species. Genes 1 and 2 are 490 and 310 bp apart, whereas gene 2 and 3 are 932 and 947-bp apart in rice and wheat, respectively. No retrotransposon insertions were observed in these regions in either species. In addition to their relative position and orientation, the predicted protein products of these three genes showed a high level of sequence identity, confirming that they are true orthologues.

In *T. monococcum*, the *Gsp-A* ^{m}*1* gene is located immediately after gene 3, but no similar gene was annotated in the corresponding region in the rice sequence (TIGR database). However, a TBLASTN comparison of *T. monococcum Gsp-A*^{m} *1* and *T. aestivum Gsp-1b* proteins with the rice BAC OSJNBb0011N16 identified a region of similarity after rice gene 3 (Fig. 3). A relatively high level of similarity was observed upon comparison of the predicted rice protein (10,321–10,446-bp) with the products of the *Gsp-A* ^{m}*1* (69%, E = 0.44) and *Gsp-D1* (67%, E = 0.12) genes (Fig. 3). A putative TATA-box was identified 525 bp downstream from the start codon and a polyadenylation signal was found 91 bp downstream of the putative stop codon. This level of protein similarity is within the range observed in other orthologous wheat/rice protein comparisons, but the E value calculated based on the size of the nr database was not significant (E = 0.12), probably because of the small size of the genes being compared. The fact that the putative rice gene is found in the same orientation and position relative to the other three colinear genes indicates that this similarity reflects the conservation of a sequence of biological importance.

To test whether a better similarity between rice and *T. monococcum Gsp-A* ^{m}*1* could be found elsewhere in the rice genome, a comparison was performed between the wheat protein and the rice genome using TBLASTN. Eleven hits were obtained with E values lower than those observed for the putative rice gene found after gene 3. Putative rice prolamin genes showed the highest similarities (E = $9e^{-11}$ to 0.075). However, the reciprocal comparison showed that this highly significant rice prolamin protein was more closely related to a wheat seed storage protein (TBLASTN, E = $8e^{-17}$) than to the Gsp proteins. This result indicates that the rice prolamin genes are most probably not orthologous to wheat *Gsp*.

Fig. 3 Alignment of *Gsp-A* ^{m}*1* with *Gsp-1b* (Accession No. X80379 from *T. aestivum*) and a putative orthologous gene on rice BAC OSJNBb0011N16. The region in the predicted rice protein that is 67% similar to the product of *Gsp-1b* is *underlined*

Rice Putative protein orthologous to *Gsp-1*

MPSFGRQAVQRVAEARQIAITLPSACNLYPTYCNIPPTAAC*

Alignment between wheat *Gsp-A*^{m}*1* and *Gsp-1b* (accession X80379 from *T. aestivum*) with the predicted rice protein

```
Expect = 0.44 Identities = 15/27 (55%), Positives = 19/27 (69%)

Rice        13   AEARQIAITLPSACNLYPTYCNIPPTA 39
                 A+  Q A +LPS CN+ P YCNIP T+
Gsp-Am1    134   AKTVQTAKSLPSKCNIDPKYCNIPITS 160


Expect = 0.12 Identities = 19/37 (51%), Positives = 25/37 (67%), Gaps = 2/37 (5%)

Rice       5   GRQAVQRVAEAR--QIAITLPSACNLYPTYCNIPPTA 39
               G + VQ+  +AR  Q A +LPS CN+ P YCNIP T+
Gsp-1b   124   GFKGVQQGLKARTVQTAKSLPSKCNIDPKYCNIPITS 160
```

The similarity observed between Gsp and the rice prolamins was probably generated by the similar pattern and spacing of cysteine residues and numerous glutamine residues, as previously reported (Anderson et al. 2001; Woo et al. 2001). Prolamins *sensu stricto* and cysteine rich proteins (i.e., puroindolines, Gsp, inhibitor of alpha-amylases) belong to the superfamily of prolamins (Shewry et al. 2002).

The rice BAC sequence between positions 10,100 and 10,500 was also investigated with different gene prediction programs. GeneMark.hmm (option 'Oryza sativa') detected an exon between positions 10,273 and 10,439-bp, but it was identified as an 'internal' axon and was in reverse orientation, showing no similarity with *Gsp-A^{m}1*. The NetPlantGene Server detected a strong coding potential between positions 10,195 and 10,450-bp in the same orientation. No rice EST showed significant similarity with the rice BAC genomic sequence between positions 10,100 and 10,500 bp.

Discussion

Distribution of genes and repetitive elements

Seven genes, including *Pina-A^{m}1*, *Pinb-A^{m}1*, and *Gsp-A^{m}1*, were identified in the 101,101-bp *T. monococcum* BAC 109N23. In addition, three hypothetical genes (hg1, 2, and 3) were annotated based on gene prediction programs, but these were not unequivocally confirmed by similarities with ESTs or with orthologous genes in other species. These hypothetical genes might be degenerated genes derived from unknown repetitive elements, and were not included in the calculation of gene density. A high gene density of 1 gene per 14 kb was found in this region. This value is well above the 1 per 230 kb that would be observed if an estimated 25,000 genes were distributed homogeneously along the *T. monococcum* genome. The observed value in BAC 109N23 is also five fold higher than the gene density reported for a 556-kb region located at fraction lengths 0.68–0.78 on the 5A^{m}L arm (Yan et al. 2003). These results confirm at the sequence level a general trend previously described at the cytogenetic level, indicating that gene densities increase from the centromeres to the telomeres (Gill et al. 1996; Akhunov et al. 2003b).

Gene distribution is not homogeneous even within BAC 109N23. Five out of the seven genes are found in the first 27 kb of the BAC, and thus this segment constitutes a high-gene-density island. In contrast, *Pina-A^{m}1* is isolated in the middle of a 69-kb region mainly represented by repetitive elements. The last gene, *Pinb-A^{m}1*, is located at the 3′ end of the BAC, and therefore it is not possible to determine if it belongs to another high-density gene island or is isolated from other genes. A similar pattern of high-gene-density islands and isolated genes, separated by repeat elements often inserted into each other, was first described in maize (SanMiguel et al. 1996) and emerges also for other wheat and barley regions analyzed so far (Wicker et al. 2001; Brooks et al. 2002; Ramakrishna et al. 2002; SanMiguel et al. 2002; Gu et al. 2003, Wicker et al. 2003b).

A comparison of the intergenic regions in this and other *T. monococcum* BACs showed the opposite trend to that described for the gene regions. Repeat sequences in BAC 109N23 represent only 42.5% of the sequence, compared with 70–80% in other *T. monococcum* BACs (Wicker et al. 2001; SanMiguel et al. 2002; Wicker et al. 2003b). In addition, the observed pattern of retrotransposon insertions was not as complex as that found elsewhere. Only two levels of nested insertions were observed in BAC 109N23, whereas more complex patterns with up to four levels of nested elements have been observed in other studies. The retrotransposon insertion pattern observed in 109N23 was consistent with their estimated insertion times. Angela_109N23-2 inserted about 1.49 Myr ago into the much older Sabrina_109N23-1, which inserted into this region some 3.10 Myr ago. These insertion times are within the same range as those calculated for other regions of the *T. monococcum* genome (SanMiguel et al. 2002) except for the Retrofit_109N23-1 element, which appears to have inserted very recently, considering that only one SNP was observed between its 196-bp LTRs.

In summary, this distal region of wheat chromosome 5A^m shows an increased gene density and a decreased proportion and complexity of repetitive elements. It is tempting to speculate that this well conserved locally gene-rich region could be the result of selective pressures that have acted to keep the organization of this region intact.

Microcolinearity between wheat and rice sequences

A good level of microcolinearity was found between rice chromosome 12 and the *Ha* locus in diploid wheat. Genes 1, 2, 3 and *Gsp* were found in the same order and orientation and at similar relative distances in the two genomes (Fig. 2). However, gene 1 was duplicated in wheat, whereas only one copy was found in rice. Two additional duplications resulted in the origin of the three related genes *Pina-A^{m}1*, *Pinb-A^{m}1*, and *Gsp-A^{m}1* in wheat after the divergence from rice. These three duplications resulted in a 75% increase (7 vs. 4) in the number of genes present in wheat relative to the orthologous region in rice. A higher frequency of duplicated genes in the Triticeae genomes relative to rice has been reported in other microcolinearity studies. Thus, Yan et al. (2003) found two gene duplications in a 550-kb region in 5A^{m}L that were absent from the orthologous region in rice.

A higher level of gene duplication in diploid wheat relative to rice was also inferred from the comparison of their RFLP maps (Dubcovsky et al. 1996). In the *T. monococcum* map, 27.7% of the cDNA clones and 34.4% of *Pst* I genomic clones detected duplicated loci, whereas less than 6% of duplicated loci were found in

the initial maps of rice (Dubcovsky et al. 1996). High levels of RFLP duplication were also observed in the large genomes of barley and polyploid wheat (Anderson et al. 1992; Kleinhofs et al. 1993; Akhunov et al. 2003b). It was recently suggested that a higher proportion of the RFLP duplications occur in the distal regions of the wheat chromosomes (Akhunov et al. 2003b). The *Ha* locus is located in the distal region of the short arm of chromosome $5A^m$, and therefore the high level of gene duplication observed in this BAC agrees with the general trend described above.

In general, colinearity among the wheat genomes is better in the proximal regions of the chromosomes than in the distal regions (Akhunov et al. 2003a). In agreement with this general trend, breaks in wheat/rice microcolinearity were frequently observed in studies involving the distal regions of the wheat genome, such as the *Lrk*/*Tak* region (Feuillet and Keller 1999), the *Sh2*/*X1*/*X2*/*A1* region (Li and Gill 2002), or the *Rpg1* region (Kilian et al. 1997). However, the *Ha* region analyzed in this study does not follow this general pattern, and shows good microcolinearity with rice (except for the duplications) in spite of its distal location on the chromosome. Another surprising aspect of this microcolinearity study was the good conservation in the relative sizes of the intergenic regions in wheat and rice. Generally, intergenic regions in wheat are considerably expanded relative to rice, but in this study a group of four genes showed similar relative intergene distances.

This level of microcolinearity facilitated the identification of a rice orthologue of the *Gsp* / *Pin* gene. This was an unexpected result because no hybridization signal was detected in previous studies between a wheat *Puroindoline-a* probe and genomic DNAs from rice, maize or sorghum (Gautier et al 2000). However, the previous result can be explained by the level of similarity found between these two short genes in this study, which is below the threshold of detection attainable under the hybridization conditions used in the previous studies. In addition, the small size of the conserved region might have also contributed to the failure to detect it by hybridization.

The small size of the predicted rice gene could be also the reason for the absence of this gene in the automatic annotation of the rice genome, and for the low expected values observed in the wheat/rice comparisons. However, the level of protein similarity found between the wheat and rice proteins (65%) is within the range of similarity values found among other orthologous genes in these two species. After 50 Myr of divergence, the non-functional regions of the Triticeae and rice genomes, including introns and intergenic regions, show no similarity (Dubcovsky et al 2001). The significance of this sequence conservation between wheat and rice is enhanced by the fact that these two regions were located in the same position relative to the other orthologous genes and in the same orientation. The putative involvement of this group of genes in disease resistance could also provide an explanation for the higher than average divergence between these genes in wheat and rice. Disease resistance genes are known to evolve faster than other genes (Michelmore and Meyers 1998). Wheat Gsp and Puroindoline proteins are more similar to each other than to the putative rice protein, suggesting that the two duplication events that originated the three wheat genes occurred after the divergence between wheat and rice.

The *T. monococcum Ha* locus contains the *Pina-1*, *Pinb-1*, and *Gsp-1* genes (similar to the D genome of soft hexaploid wheat varieties, Gautier et al. 1994; Sourdille et al. 1996). These three genes have also been observed in other diploid wheats, but the *Pina-1* and *Pinb-1* genes are absent from the A and B genomes of polyploid wheats (Gautier et al. 2000). Additional instances of deletion of the *Pina-D1* gene or of the complete *Pina-D1* - *Pinb-D1* - *Gsp-D1* region have also been reported in hexaploid wheat (Tranquilli et al. 2002). The presence of large regions of highly repetitive elements around the *Pina-1* and *Pinb-1* genes might provide an explanation for the frequent occurrence of deletions that affect these two genes (e.g. resulting from unequal crossing over between repeat elements). Comparison of the complete *Ha* locus described here with the sequences of the orthologous region from the A and B genomes of tetraploid and hexaploid wheat will provide a better understanding of the mechanisms involved in the elimination of the puroindoline genes from the latter genomes.

Acknowledgements This research was supported by USDA-IFAFS Competitive Grant No. 2001-04462. The authors thank Marie-Françoise Gautier for her advice on the prolamin family. They also thank Philippe Leroy and Bastien Laubin for their contributions to the sequence annotation and Gaëtan Droc for his help with bioinformatics

References

Akhunov ED, et al (2003a) Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates. Proc Natl Acad Sci USA 100:10836–10841

Akhunov ED, et al (2003b) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res 13:753–763

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

Anderson JA, Ogihara Y, Sorrells ME, Tanksley SD (1992) Development of a chromosomal arm map for wheat based on RFLP markers. Theor Appl Genet 83:1035–1043

Anderson OD, Hsia CC, Adalsteins AE, Lew EJ-L, Kasarda DD (2001) Identification of several new classes of low-molecular-weight wheat gliadin-related proteins and genes. Theor Appl Genet 103:307–315

Anjum FM, Walker CE (1991) Review on the significance of starch and protein to wheat kernel hardness. J Sci Food Agric 56:1–13

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Beecher B, Bettge A, Smidansky E, Giroux J (2002) Expression of wild-type *pinB* sequence in transgenic wheat complements a hard phenotype. Theor Appl Genet 105:870–877

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

Bennetzen JL, Ramakrishna W (2002) Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. Plant Mol Biol 48:821–827

Brooks SA, Huang L, Gill BS, Fellers JP (2002) Analysis of 106 kb of contiguous DNA sequence from the D genome of wheat reveals high gene density and a complex arrangement of genes related to disease resistance. Genome 45:963–972

Brunner S, Keller B, Feuillet C (2003) A large rearrangement involving genes and low-copy DNA interrupts the microcollinearity between rice and barley at the *Rph7* locus. Genetics 164:673–683

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268:78–94

Devos KM, Brown JK, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. Genome Res 12:1075–1079

Dubcovsky J, Luo MC, Zhong GY, Bransteitter R, Desai A, Kilian A, Kleinhofs A, Dvorak J (1996) Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L. Genetics 143:983–999

Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan L, Shiloff BA, Bennetzen JL (2001) Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol 125:1342–1353

Echenique V, Stamova B, Wolters P, Lazo G, Carollo L, Dubcovsky J (2002) Frequencies of *Ty1-copia* and *Ty3-gypsy* retroelements within the Triticeae EST databases. Theor Appl Genet 104:840–844

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8:186–194

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

Fu H, Dooner HK (2002) Intraspecific violation of genetic colinearity and its implications in maize. Proc Natl Acad Sci USA 99:9573–9578

Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA 93:10274–10279

Gautier MF, Aleman ME, Guirao A, Marion D, Joudrier P (1994) *Triticum aestivum* puroindolines, two basic cysteine-rich seed proteins: cDNA sequence analysis and developmental gene expression. Plant Mol Biol 25:43–57

Gautier MF, Cosson P, Guirao A, Alary R, Joudrier P (2000) Puroindoline genes are highly conserved in diploid ancestor wheats and related species but absent in tetraploid *Triticum* species. Plant Sci 153:81–91

Gill KS, Gill BS, Endo T, Boyko EV (1996) Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. Genetics 143:1001–1012

Giroux MJ, Morris CF (1998) Wheat grain hardness results from highly conserved mutations in the friabilin components puroindoline a and b. Proc Natl Acad Sci USA 95:6262–6266

Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. Genome Res 8:195–202

Greenwell P, Schofield JD (1986) A starch granule protein associated with endosperm softness in wheat. Cereal Chem 63:379–380

Gu YQ, Anderson OD, Londeore C, Kong X, Chibbar RN, Lazo GR, Wilson C (2003) Structural organization of the barley *D-hordein* locus in comparison with its orthologous regions of wheat genomes. Genome 46:1084–1097

Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S (1996) Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. Nucleic Acids Res 24:3439–3452

Jurka J (2000) Repbase Update: a database and an electronic journal of repetitive elements. Trends Genet 9:418–420

Keller B, Feuillet C (2000) Colinearity and gene density in grass genomes. Trends Plant Sci 5:1360–1251

Kellogg EA (2001) Evolutionary history of the grasses. Plant Physiol 125:1198–205

Kilian A, Chen J, Han F, Steffenson B, Kleinhofs A (1997) Towards map-based cloning of the barley stem rust resistance genes *Rpg1* and *rpg4* using rice as an intergenomic cloning vehicle. Plant Mol Biol 35:187–195

Kleinhofs A, et al (1993) A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. Theor App Genet 86:705–712

Krishnamurthy K, Balconi C, Sherwood JE, Giroux MJ (2001) Wheat puroindolines enhance fungal disease resistance in transgenic rice. Mol Plant Microbe Interact 14:1255–1260

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software. Bioinformatics 17:1244–1245

Law CN, Young CF, Brown JWS, Snape JW, Worland AJ (1978) The study of grain protein control in wheat using whole chromosome substitution lines. In: Seed protein improvement by nuclear techniques. International Atomic Energy Agency, Vienna, Austria, pp 483–502

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

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

Lukashin AV, Borodovsky M (1998) GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res 26:1107–15

Mattern PJ, Morris R, Schmidt JW, Johnson VA (1973) Location of genes for kernel properties in the wheat cultivar Cheyenne using chromosome substitution lines. Proc 4th Int Wheat Genet. Symp, University of Missouri, Columbia, Mo., pp 703–707

McCarthy E, McDonald J (2003) LTR_STRUC: a novel search and identification program for LTR retrotransposons. Bioinformatics 19:362–367

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113–1130

Moore G, Devos KM, Wang Z, Gale MD (1995) Cereal genome evolution. Grasses, line up and form a circle. Curr Biol 5:737–739

Morris CF (2002) Puroindolines: the molecular genetic basis of wheat grain hardness. Plant Mol Biol 48:633–647

Panstruga R, Büschges R, Piffanelli P, Schulze-Lefert P (1998) A contiguous 60-kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucleic Acids Res 26:1056–1062

Rahman S, Jolly JC, Skerritt JH, Wallosheck A (1994) Cloning of a wheat 15-kDa grain softness protein (GSP). GSP is a mixture of puroindoline-like polypeptides. Eur J Biochem 223:917–925

Ramakrishna W, Dubcovsky J, Park YJ, Busso C, Emberton J, SanMiguel P, Bennetzen JL (2002) Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. Genetics 162:1389–1400

Rostoks N, Park YJ, Ramakrishna W, Ma J, Druka A, Shiloff BA, SanMiguel PJ, Jiang Z, Brueggeman R, Sandhu D, Gill K, Bennetzen JL, Kleinhofs A (2002) Genomic sequencing reveals gene content, genomic organization, and recombination relationships in barley. Funct Integr Genomics 2:51–59

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested restrotransposons in the intergenic regions of the maize genome. Science 274:765–768

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

Shewry PR, Beaudoin F, Jenkins J, Griffiths-Jones S, Mills EN (2002) Plant protein families and their relationships to food allergy. Biochem Soc Trans 30:906–910

Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 10:908–915

Sourdille P, Perretant MR, Charmet G, Leroy P, Gautier MF, Joudrier P, Nelson JC, Sorrells ME, Bernard M (1996) Linkage between RFLP markers and genes affecting kernel hardness in wheat. Theor Appl Genet 93:580–586

Symes KJ (1965) The inheritance of grain hardness in wheat as measured by the particle size index. Aust J Agric Res 20:971–979

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

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Tranquilli G, Lijavetzky D, Muzzi G, Dubcovsky J (1999) Genetic and physical characterization of grain texture-related loci in diploid wheat. Mol Gen Genet 262:846–850

Tranquilli G, Heaton J, Chicaiza O, Dubcovsky J (2002) Substitutions and deletions of genes related to grain hardness in wheat and their effect on grain texture. Crop Sci 42:1812–1817

Turnbull KM, Turner M, Mukai Y, Yamamoto M, Morell MK, Appels R, Rahman S (2003) The organization of genes tightly linked to the *Ha* locus in *Aegilops tauschii*, the D-genome donor to wheat. Genome 46:330–338

Van Deynze AE, Dubcovsky J, K.S. G, Nelson JC, Sorrells ME, Dvorak 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

Wicker T, Stein N, Albar L, Feuillet C, Schlagenheuf E, Keller B (2001) Analysis of a contiguous 211-kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 26:307–316

Wicker T, Guyot R, Yahiaoui N, Keller B (2003a) CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements. Plant Physiol 132:52–63

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, Dubcovsky J, Keller B (2003b) Rapid genome divergence at orthologous Low Molecular Weight Glutenin loci of the A and A(m) genomes of wheat. Plant Cell 15:1186–1197

Woo YM, Hu DW, Larkins BA, Jung R (2001) Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. Plant Cell 13:2297–2317

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