Funct Integr Genomics (2004) 4: 102–117
DOI 10.1007/s10142-004-0109-y

ORIGINAL PAPER

Nagendra K. Singh · Saurabh Raghuvanshi · Subodh K. Srivastava · Anupama Gaur · Ajit K. Pal · Vivek Dalal · Archana Singh · Irfan A. Ghazi · Ashutosh Bhargav · Mahavir Yadav · Anupam Dixit · Kamlesh Batra · Kishor Gaikwad · Tilak R. Sharma · Amitabh Mohanty · Arvind K. Bharti · Anita Kapur · Vikrant Gupta · Dibyendu Kumar · Shubha Vij · Ravi Vydianathan · Parul Khurana · Sulabha Sharma · W. Richard McCombie · Joachim Messing · Rod Wing · Takuji Sasaki · Paramjit Khurana · Trilochan Mohapatra · Jitendra P. Khurana · Akhilesh K. Tyagi

Sequence analysis of the long arm of rice chromosome 11 for rice–wheat synteny

Received: 20 August 2003 / Revised: 5 February 2004 / Accepted: 21 February 2004 / Published online: 14 April 2004

N. K. Singh · S. K. Srivastava · A. K. Pal · V. Dalal · A. Singh · I. A. Ghazi · A. Bhargav · M. Yadav · A. Dixit · K. Batra · K. Gaikwad · T. R. Sharma · T. Mohapatra
Indian Initiative for Rice Genome Sequencing, National Research Centre on Plant Biotechnology, Indian Agricultural Research Institute,
New Delhi, 110012, India

S. Raghuvanshi · A. Gaur · A. Mohanty · A. K. Bharti · A. Kapur · V. Gupta · D. Kumar · S. Vij · R. Vydianathan · P. Khurana · S. Sharma · P. Khurana · J. P. Khurana · A. K. Tyagi (✉)
Indian Initiative for Rice Genome Sequencing, Department of Plant Molecular Biology, University of Delhi South Campus,
New Delhi, 110021, India
e-mail: akhilesh@genomeindia.org
Tel.: +91-11-24673216
Fax: +91-11-24109450

W. R. McCombie
Genome Research Centre, Cold Spring Harbor Laboratory,
1 Bungtown Road,
New York, USA

J. Messing
The Plant Genome Initiative at Rutgers–Waksman Institute,
Rutgers University,
190 Frelinghuysen Road,
Piscataway, NJ, 08873, USA

R. Wing
Department of Plant Sciences, 303 Forbes Building, Arizona Genomics Institute, The University of Arizona,
Tucson, AZ, 85721, USA

T. Sasaki
Department of Genome Research, National Institute of Agrobiological Sciences,
1-2, Kannondai 2-chome,
Tsukuba, Ibaraki, 305-8602, Japan

Abstract The DNA sequence of 106 BAC/PAC clones in the minimum tiling path (MTP) of the long arm of rice chromosome 11, between map positions 57.3 and 116.2 cM, has been assembled to phase 2 or PLN level. This region has been sequenced to 10× redundancy by the Indian Initiative for Rice Genome Sequencing (IIRGS) and is now publicly available in GenBank. The region, excluding overlaps, has been predicted to contain 2,932 genes using different software. A gene-by-gene BLASTN search of the NCBI wheat EST database of over 420,000 cDNA sequences revealed that 1,143 of the predicted rice genes (38.9%) have significant homology to wheat ESTs (bit score ≥ 100). Further BLASTN search of these 1,143 rice genes with the GrainGenes database of sequence contigs containing bin-mapped wheat ESTs allowed 113 of the genes to be placed in bins located on wheat chromosomes of different homoeologous groups. The largest number of genes, about one-third, mapped to the homoeologous group 4 chromosomes of wheat, suggesting a common evolutionary origin. The remaining genes were located on wheat chromosomes of different groups with significantly higher numbers for groups 3 and 5. Location of bin-mapped wheat contigs to chromosomes of all the seven homoeologous groups can be ascribed to movement of genes (transpositions) or chromosome segments (translocations) within rice or the hexaploid wheat genomes. Alternatively, it could be due to ancient duplications in the common ancestral genome of wheat and rice followed by selective elimination of genes in the wheat and rice genomes. While there exists definite conservation of gene sequences and the ancestral chromosomal identity between rice and wheat, there is no obvious conservation of the gene order at this level of resolution. Lack of extensive

colinearity between rice and wheat genomes suggests that there have been many insertions, deletions, duplications and translocations that make the synteny comparisons much more complicated than earlier thought. However, enhanced resolution of comparative sequence analysis may reveal smaller conserved regions of colinearity, which will facilitate selection of markers for saturation mapping and sequencing of the gene-rich regions of the wheat genome.

Keywords Rice · Wheat · Synteny · Plant genome

Introduction

Comparative studies of plant genomes have provided some of the most useful knowledge about the evolutionary biology of crop plants (Heslop-Harrison 2000). This understanding will help devise potent new tools for the genetic improvement of such species for increasing food production and quality, particularly in the members of the grass family Poaceae, which provides more than 60% of the world's food supply. Early in the last century, Kihara (1924) and Sears (1941) did pioneering work on the genome analysis of wheat establishing very clearly the presence of polyploidy and seven homoeologous groups of chromosomes originating from a common ancestral genome. The gene sets of homoeologous chromosomes of rye and barley are so well conserved that they can substitute the loss of complete wheat chromosomes of roughly twice the size of the entire rice genome in intergeneric chromosomal substitution lines (Islam and Shepherd 1992). Molecular genetic studies based on sets of cloned DNA probes hybridizing to genomic DNAs of different species have enlarged our understanding of the conservation of gene synteny within the grass family, leading to the concept of alignment of different grass genomes at the macro-level (Ahn et al. 1993; Bennetzen and Freeling 1993; Gale and Devos 1998; Paterson et al. 2000). Sequence-based comparative genomics for evolutionary studies and for transferring information from completely sequenced model species, e.g., rice, to related large-genome crop species holds promise to revolutionize functional genomics and molecular breeding strategies for crop improvement (Appels et al. 2003; Bowers et al. 2003; Vandepoele et al. 2003). Earlier investigations based on a small number of RFLP markers have shown a variable level of colinearity between rice and wheat (Ahn et al. 1993; Kurata et al. 1994; Deynze et al. 1995; Gale and Devos 1998). Recently, a genome-wide comparison of 4,485 wheat expressed sequence tags (ESTs) physically mapped to wheat chromosome bins (Faris et al. 2002; Qi et al. 2003) with 2,251 BAC/PAC clones ordered on rice chromosomes has revealed a much more complex nature of the comparative genome relationship (Sorrells et al. 2003). In the present study, we report on the outcome of a gene-by-gene comparison of the sequence homology of 2,932 predicted rice genes from the long arm of chromosome 11 with wheat ESTs and their location on wheat chromosome bins.

Materials and methods

The present comparative study was performed on a ~12 Mb region from the long arm of rice chromosome 11 sequenced by the Indian Initiative for Rice Genome Sequencing (IIRGS) from 106 ordered BAC/PAC clones in the minimum tiling path of this region (see http://www.genomeindia.org/). Gene prediction was carried out primarily using RiceGAAS (Sakata et al. 2002; http://ricegaas.dna.affrc.go.jp/). The genes predicted by AutoPredgeneset of the RiceGAAS were taken into consideration for further analysis. Wherever RiceGAAS data were not available, the genes were predicted by FGENESH trained for monocot plant species (http://www.softberry.com/berry.phtml). Subsequently, the genomic sequence of each of the 2,932 predicted genes in the non-overlapping regions of the BAC/PAC sequences were downloaded to create a local database for further analysis. Experiments were designed to optimize the settings of BLASTN search options and reporting parameters in order to detect even a distant homology between the rice genes and wheat cDNA sequences. This was necessary because the default search parameters, being highly stringent, were picking up only short sequence stretches of high homology. By taking known rice genes, two each with high homology (adenosyl homocysteinase and hydroxymethyl transferase), medium homology (acyl CoA synthase and flavonone-3-sulphotransferase) and low homology (glutelin I precursor and 22 kDa kafirin type protein) with wheat, 96 different combinations of search parameters were explored. The optional search parameters included all permutations of expect values e (10, 100 and 1,000), low complexity filter for repeats (with and without), open gap cost G (2 and 5), extend gap cost E (1 and 2), penalty for a mismatch q (−1 and −3) and reward for a match r (1 and 5). The search results with the above six genes using 96 different optional parameter settings were compared to derive the optimum settings. Finally, an optimized value of G =5, E =1, q =−1, r =1, a word size of 11 and expect value of 10, without the low complexity filter, were used for further analysis in order to record the optimum number of matches between rice genes and wheat ESTs. Sets of individual genes, predicted from a BAC/PAC clone, were searched using the MEGABLAST tool against the wheat EST database of NCBI (http://www.ncbi.nlm.nih.gov/dbEST/index.html). Details of only top hits of the BLASTN search results with each rice gene were tabulated in an Excel file and all those genes showing a bit score of 100 or more were extracted into a separate file. The sequences of these matching rice genes were then searched against the GrainGenes database of wheat sequence contigs containing bin-mapped ESTs to find their physical location on the wheat chromosomes (http://www.graingenes.org/cgi-bin/ace/search/graingenes). Once again, the optimized BLASTN search parameters, as mentioned above, were used to find the bin-mapped wheat homologues. The bit scores of real matches with the bin-mapped wheat ESTs were usually higher than the initial bit score obtained with the NCBI ESTs, and the absence of matches were indicated by a very low bit score of around 50. The matching rice genes were then plotted for their location to the specific bins (segments) of wheat chromosomes (Qi et al. 2003). The supplemental data resulting from such an analysis were tabulated and are available at http://www.genomeindia.org/rice-wheat-synteny and also from the F&IG web site.

Results and discussion

Optimization of the BLASTN search parameters

Rice is the first genome of an economically important plant species to be completely sequenced. A high quality

phase 2 sequence with tenfold coverage has already been generated by the IRGSP, a consortium of publicly funded laboratories from ten countries, and is available at the RGP website http://rgp.dna.affrc.go.jp/. This information is playing a pivotal role in gene discovery and comparative genome analysis of important cereal crops, including wheat, barley, maize, sorghum, sugarcane and millets (Sorrells et al. 2003; Vandepoele et al. 2003). The outcome of such an analysis will be largely dependent on genome-wide rearrangements and the rate of genome evolution, which could have influenced the gene order along the chromosomes (Bowers et al. 2003). Furthermore, the results of such comparative analyses will also depend on the types of databases and computational tools employed. There is a large EST database resource of about half a million entries for wheat, which is next only to human, mouse and rat EST databases (http://www.ncbi.nlm.nih.gov/dbEST/index.html). More than 7,533 of these ESTs have now been mapped to specific wheat chromosome bins using segmental deletion lines and a public database has been created (http://wheat.pw.usda.gov/NSF/progress_mapping.html). These two databases were used as a valuable resource for comparison of predicted rice genes with the wheat ESTs.

In an effort to optimize the BLASTN search parameters, it was observed that the default search parameters of NCBI, i.e., use of low complexity filter, open gap cost (G) of 5, extend gap cost (E) of 2, penalty for a mismatch (q) of −3 and reward for a match (r) of 1, were actually too stringent for comparing distantly related genes and favored outputs of only short stretches with high percent homologies. It was interesting to note that with the NCBI default search parameters, two seed storage protein gene homologues, namely rice glutelin precursor and 22-kDa kafirin, gave significantly lower bit scores of 84 and 44, respectively, and would not have registered as wheat homologues with the cut off bit score of 100. Hybridization data have previously assumed that 22-kDa kafirin genes are completely absent in rice and orthologous sequence alignments between maize, sorghum and rice have shown that orthologous kafirin sequences at rice chromosome 11 position 97.3 cM are missing (Song et al. 2002). However, with the optimized search parameters of reducing the cost of extending a gap to 1 and reducing the penalty for a mismatch to −1, the bit scores for both of these matches increased to more than 200, which is indeed significant. For highly conserved genes having more than 90% sequence homology, e.g., adenosyl homocysteinase and hydroxymethyl transferase, the relaxing of search parameters did not make any significant difference.

However, extreme relaxation of the search parameters resulted in loss of discrimination between genes of high, medium and low levels of sequence conservation with respect to their percent homology with wheat ESTs. For example, with the use of the most relaxed search parameters of G =2, E =1, q =−3 and r =5, the bit scores for the set of three genes with high, medium and low sequence conservation, namely hydroxymethyl transferase, acyl CoA synthase and 22-kDa kafirin, were each increased to above 400 but the percent sequence homology dropped to 59%, 47% and 49%, compared to 91%, 85% and 64%, respectively, obtained with the optimized search parameters (data not shown). Thus, it is not desirable to use either the highly stringent default BLASTN search parameters of the NCBI or highly relaxed search parameters described above. Hence, the optimized search parameters of G =5, E =1, q =−1, r =1, word size 11 and expected value of 10, without the low complexity filter, have been used throughout this study.

Analysis of gene conservation between rice and wheat

The aim of the present study was to investigate the extent of conservation of genes in terms of sequence homology and their physical location on the rice and wheat chromosomes, taking into account the ~12 Mb region from the long arm of rice chromosome 11 sequenced by the IIRGS as a test case. Thus, an in silico gene prediction was performed with this region of rice chromosome 11, represented by 106 ordered BAC/PAC clones, and a local database created for further analysis. In a recent report on the comparative analysis of wheat and rice genomes, only 29 BAC/PAC clones were available for the whole of rice chromosome 11, since only the data up to May 2002 were accessed (Sorrells et al. 2003). Out of the 106 BAC/PAC

Table 1 Classification of genes from rice chromosome 11 (57.3–116.2 cM) identified using AutoPredgeneset (RiceGAAS) and FGENESH gene prediction systems. Genes showing significant homology to wheat ESTs have been tabulated

Gene category based on product	No. of predicted rice genes (%)	No. of rice genes matching with wheat ESTs (%)	No. of rice genes matching with bin-mapped wheat ESTs (%)
Exact/putative[a]	733 (25)	492 (43)	75 (66)
Unknown[b]	464 (16)	89 (8)	27 (24)
Hypothetical[c]	1,735 (59)	562 (49)[d]	11 (10)[d]
Total	2,932 (100)	1,143 (100)	113 (100)

[a] 100% homologous all over the length of cDNA and protein/high degree of homology (bit score >100 and e-value < 10^{-20}) at both DNA and protein level

[b] High degree of homology with EST (bit score >100 and e-value < 10^{-20}) but no protein homology

[c] Predicted genes with no significant homology hit in either the EST or protein database

[d] It may be noted that some rice genes predicted to be hypothetical showed homology to wheat ESTs on further analysis

clones analyzed in the present study, only a few are common to the study by Sorrells et al. (2003). A total of 2,932 genes were predicted in the non-overlapping regions of these 106 clones and, depending upon their homology to a known protein or EST, were classified as exact/putative, unknown or hypothetical genes (Table 1). The first round of BLASTN search was performed with the genomic sequences of the 2,932 predicted rice genes against the EST database of NCBI by limiting the search by Entrez query to *Triticum aestivum* (ORGN). Hits with bit score values of 100 or more were treated as significant. Overall, 1,143 of the predicted rice genes (38.9%) showed significant homology with the wheat ESTs (Table 1). This percentage of conserved genes between rice and wheat is similar to that reported by Sorrells et al. (2003) who found that 37% of the all bin-mapped wheat ESTs matched with the sequences of rice BAC clones. It is significant that 61% of the predicted rice genes have no significant homology with the wheat ESTs and hence these are either lineage-specific rice genes or could represent truly hypothetical genes/rarely transcribed genes. Also, it is important to note that a large number of genes (562 genes) classified under the hypothetical category by the RiceGAAS are matching significantly with the wheat ESTs and, therefore, these should be regrouped into the unknown category (Table 1).

All the 1,143 genes showing homology with wheat ESTs were then taken for a second round of BLASTN search against the GrainGenes database of wheat contigs containing bin-mapped ESTs (Table 2; supplemental data at http://www.genomeindia.org/rice-wheat-synteny and F&IG web site). The genes in Table 2 have been arranged according to the pre-established MTP in the 57.3 to 116.2 cM region of the rice chromosome 11 and thus essentially represent the physical order in which they are present in rice chromosome 11 (Chen et al. 2002). However, it should be noted that the order of BAC/PAC clones in Table 2 of this article as shown at http://www.genomeindia.org/rice-wheat-synteny (and F&IG web site) cannot be directly compared to the sequencing status table on the web site http://www.genomeindia.org/seq_status.htm, because the latter is a dynamic table and has many additional clones. In contrast to the search results with the NCBI wheat EST database, only 113 (9.8%) of the 1,143 genes showed significant matches with the bin-mapped EST contigs indicating that only a fraction of the homologous wheat ESTs have been bin-mapped. Nevertheless, this number is much higher than 11 matches for the entire rice chromosome 11 reported earlier by Sorrells et al. (2003), mainly due to the availability of additional rice genome sequence information for the present study. The 113 identified rice genes were homologous to a total of 83 wheat cDNA contigs, as several of the rice genes were in two or more copies due to tandem duplications. Thus, out of the 113 rice genes 60 were in single copies, 19 in two copies, 2 in three copies and 1 each in four copies and five copies. This shows that the tandem duplication of genes along the chromosome is a common phenomenon but occurrence of genes with three or more copies is rare, at least in this part of the rice genome. Genome-wide tandem duplication of genes is also a striking feature of the *Arabidopsis* genome, in addition to their duplication on other chromosomes (Bowers et al. 2003). About two-thirds of the rice genes matching bin-mapped wheat ESTs belonged to the exact/putative category, whereas 24% and 10% of them were of unknown and hypothetical categories, respectively (Table 1). This may just represent a bias towards genes with known functions in the wheat bin-mapping studies. The genes in the exact/putative category represent several physiologically important proteins such as auxin response factor, chalcone synthase, polygalacturonidase, GTP-binding protein, sugar transporter, cytochrome P450, pyruvate decarboxylase, pathogenesis-related proteins, lipoxygenase, flavonol glucosyltransferase, reverse transcriptase and so on. Detailed information regarding the sequences of all the 113 rice genes matching with bin-mapped wheat ESTs can be obtained from the supplemental material provided at the web link http://www.genomeindia.org/rice-wheat-synteny and the F&IG web site.

Some of the 113 rice genes showed significant homology to more than one wheat EST contig but the match with the highest bit score was chosen for simplicity of the analysis. A pictorial view of the distribution of these 113 rice genes on the 21 wheat chromosomes belonging to the seven homoeologous groups of the A, B and D genomes is given in Fig. 1. Color fillings in the individual cells depict presence of a mapped-wheat EST contig in relation to the region of the rice chromosome 11 from 57.3 to 116.2 cM. There are several interesting features of the comparison shown in Fig. 1, the most important one being that the majority of the rice genes from this region of rice chromosome 11 map to the homoeologous group 4 of wheat, suggesting that rice chromosome 11 and wheat group 4 chromosomes share a common origin. Sorrells et al. (2003) have recently shown the relationship between 12 rice chromosomes and 7 homoeologous groups of wheat based on a large number of reference points (i.e., 1,613 matching bin-mapped wheat ESTs), but information about chromosomes 9, 11 and 12 was rather limited. Other studies with rice chromosome 11 have shown it to be syntenic to wheat chromosomes of group 3 (Ahn et al. 1993) and group 6 (Deynze et al. 1995; Gale and Devos 1998) but these conclusions were based on a limited number of RFLP marker sets. Although group 4 chromosomes of wheat are reported to be syntenic to chromosome 3 of rice in the study of Sorrells et al. (2003), this is not inconsistent with our results as only rice chromosomes 1 and 2 are related to single wheat chromosomes (group 3 and group 6, respectively). The remaining ten chromosomes of rice are related to five chromosome groups of wheat in a way that for every wheat chromosome there are two rice chromosomes. Thus, based on Sorrells et al. (2003) and the present work rice chromosomes 3 and 11 are related to group 4 chromosomes of wheat, rice chromosomes 4 and 7 to group 2 of wheat, rice chromosomes 5 and 10 to group 1 of wheat, rice

Table 2 List of rice genes showing significant homology with wheat contigs containing bin-mapped ESTs. The genes are arranged according to the established minimum tiling path on the long arm of rice chromosome 11 (57.3–116.2 cM)

S. no.	Rice genes information				BLAST search results			Mapped wheat EST	
	BAC/PAC clone	ID	Predicted cDNA size (bp)	Product[a]	Score	*e*-value	% Homology	Contig	Chromosome bin
1	OSJNBa0041C22	16	1,380	Hydroxymethyl-transferase	1,386	0.0	975/1,071 (91%)	11244	C-2AL1-0.85
2	OSJNBa0041C22	24	387	Hypothetical	239	1×10^{-63}	173/196 (88%)	9969	6BS5-0.76-1.05 5AL23-0.87-1.00 4DL12-0.71-1.00
3	OSJNBa0070D14	33	2,331	TNP-2 like transposon	198	2×10^{-50}	239/339 (70%)	7888	3DL3-0.81-1.00 C-3BL2-0.22
4	OSJNBb0006N14	21	2,199	F-box protein	139	8×10^{-33}	176/260 (67%)	1812	7AS1-0.89-1.00
5	OSJNBb0091D07	26	1,481	Cyt. P450	271	2×10^{-72}	362/539 (67%)	7298	3AS4-0.45-1.00 3BS9-0.57-0.78 3DS6-0.55-1.00
6	OSJNBb0057M06	11	1,395	Cyt. P450	413	10^{-115}	499/713 (69%)	7298	3AS4-0.45-1.00 3BS9-0.57-0.78 3DS6-0.55-1.00
7	OSJNBa0027P09	32	513	Salt inducible protein	450	10^{-127}	399/505 (79%)	9048	7DS4-0.61-1.00
8	OSJNBa0078F20	17	1,671	Cyt. P450	131	1×10^{-30}	255/418 (61%)	8719	3BS9-0.57-0.78 3DS6-0.55-1.00 3AS4-0.45-1.00
9	OSJNBa0057E15	09	1,671	Hypothetical	131	1×10^{-30}	255/418 (61%)	8719	3BS9-0.57-0.78 3DS6-0.55-1.00 3AS4-0.45-1.00
10	OSJNBa0057E15	26	1,317	Enolase phosphatase	237	1×10^{-62}	180/211 (85%)	5101	5BS4-0.43-0.56 5AS1-0.40-0.75 C-5DS1-0.63
11	OSJNBa0057E15	27	315	Hypothetical	228	2×10^{-60}	177/211 (83%)	5101	5BS4-0.43-0.56 5AS1-0.40-0.75 C-5DS1-0.63
12	OSJNBa0057E15	29	2,982	Dihydrofolate reductase	641	0.0	836/1,248 (66%)	7791	7AS5-0.59-0.89 3DL3-0.81-1.00 7DS4-0.61-1.00 3AL5-0.78-1.00 4AL4-0.80-1.00
13	OSJNBb0041J18	01	1,035	6-phosphoglucodehydro-genase	502	10^{-142}	675/1014 (66%)	7791	7AS5-0.59-0.89 3DL3-0.81-1.00 7DS4-0.61-1.00 3AL5-0.78-1.00 4AL4-0.80-1.00
14	OSJNBb0041J18	29	651	Hypothetical	171	6×10^{-43}	147/182 (80%)	9887	3AS4-0.45-1.00 1BS.sat18-0.50-1.00 3BS8-0.78-1.00 1AS3-0.86-1.00 1DS5-0.70-1.00
15	OSJNBb0085G12	22	2,397	FAD oxidoreductase	190	5×10^{-48}	340/541 (62%)	3591	1AL3-0.61-1.00 1BL3-0.85-1.00 2DS5-0.47-1.00
16	OSJNBb0085G12	27	210	Hypothetical	264	1×10^{-71}	187/208 (89%)	5339	5AL23-0.87-1.00 4BL5-0.86-1.00 4DL13-0.56-0.71

Table 2 (continued)

S. no.	Rice genes information				BLAST search results			Mapped wheat EST	
	BAC/PAC clone	ID	Predicted cDNA size (bp)	Product[a]	Score	*e*-value	% Homology	Contig	Chromosome bin
17	OSJNBb0085G12	30	681	Hypothetical	125	5×10^{-29}	165/242 (68%)	8629	C-3AL3-0.42 C-3BL2-0.22 3DL2-0.27-0.81
18	OSJNBb0018F17	11	1,074	Ubiquitin activator	651	0.0	465/520 (89%)	8531	C7B
19	OSJNBb0008H02	17	912	Nodulin MTn3 protein	643	0.0	501/587 (85%)	1425	C-2AL1-0.85
20	OSJNBa0002A19	09	1,563	Serine palmitoyl transferase	470	10^{-132}	414/527 (78%)	7178	3DL2-0.27-0.81 3BL7-0.63-1.00
21	OSJNBb0001I22	16	339	Hypothetical	123	7×10^{-29}	167/247 (67%)	11107	6B 6D C-7BS1-0.27
22	P0410D09	05	2,436	Auxin response factor	313	3×10^{-85}	348/484 (71%)	8017	3BL7-0.63-1.00 3AL5-0.78-1.00
23	OSJNBb0056I10	03	2,445	Auxin response factor	191	2×10^{-48}	168/216 (77%)	8017	3BL7-0.63-1.00 3AL5-0.78-1.00
24	OSJNBb0056I10	23	1,227	Cyt. P450	935	0.0	687/775 (88%)	11126	4DS1-0.53-0.67 4AL12-0.43-0.59 4BS4-0.37-0.57
25	OSJNBa0078F08	14	1,479	Hypothetical	171	2×10^{-42}	674/1201 (56%)	9288	C-5AL10-0.57[a] C-5BL14-0.75[a]
26	OSJNBa0078F08	24	561	Unknown	497	10^{-141}	455/567 (80%)	9944	C-2BS1-0.53 2DS1-0.33-0.47 2AS5-0.78-1.00
27	OSJNBa0078F08	28	1,329	Auxin response factor	148	6×10^{-36}	237/376 (63%)	802	3A 3BL2-0.22-0.50
28	OSJNBa0078F08	32	1,008	Chalcone synthase	122	6×10^{-28}	136/196 (69%)	11674	5DS2-0.78-1.00
29	OSJNBa0078F08	35	1,242	Chalcone synthase	320	2×10^{-87}	400/589 (67%)	5170	1DS5-0.70-1.00 1AS3-0.86-1.00
30	OSJNBa0078F08	38	1,206	Chalcone synthase	299	3×10^{-81}	400/592 (67%)	5170	1BS.sat19-0.31-0.50 1DS5-0.70-1.00 1AS3-0.86-1.00
31	OSJNBa0078F08	39	1,314	Chalcone synthase	378	10^{-105}	427/601 (71%)	5170	1BS.sat19-0.31-0.50 1DS5-0.70-1.00 1AS3-0.86-1.00
32	OSJNBa0033H01	05	1,197	Chalcone synthase	278	4×10^{-75}	416/652 (63%)	11674	5DS2-0.78-1.00
33	OSJNBa0080I08	27	2,946	GP-91 protein	283	4×10^{-76}	342/501 (68%)	5610	3DL2-0.27-0.81 3BL10-0.50-0.63 3AL3-0.42-0.78
34	OSJNBa0041J17	08	2,667	Respiratory burst oxidase	302	7×10^{-82}	305/420 (72%)	5966	5AL23-0.87-1.00
35	OSJNBa0041J17	20	840	Citrate synthase	499	10^{-141}	352/390 (90%)	10911	C-6DL6-0.29 6DS2-0.45-0.79
36	OSJNBa0041J17	24	984	Endoxyloglucan transferase	380	10^{-105}	275/311 (88%)	2228	C-4AL12-0.43
37	OSJNBa0041J17	29	1,128	O-methyl transferase	524	10^{-149}	610/875 (69%)	7329	6DS6-0.99-1.00
38	OSJNBa0007D07	02	888	Endoxyloglucan transferase	380	10^{-106}	275/311 (88%)	2228	C-4AL12-0.43
39	OSJNBb0005H02	09	567	Hypothetical	177	7×10^{-45}	190/254 (74%)	1189	4DL9-0.31-0.56 4BL5-0.86-1.00
40	OSJNBb0005H02	10	1,083	GF14-d protein	624	10^{-179}	572/746 (76%)	11684	4DS3-0.67-0.82 4AS4-0.63-0.76

Table 2 (continued)

S. no.	Rice genes information				BLAST search results			Mapped wheat EST	
	BAC/PAC clone	ID	Predicted cDNA size (bp)	Product[a]	Score	*e*-value	% Homology	Contig	Chromosome bin
41	OSJNBa0063F24	03	735	Purple acid phosphatase	424	10^{-119}	396/525 (75%)	7253	4DL12-0.71-1.00 C-4AL12-0.43 4BL1-0.71-1.00[a]
42	OSJNBa0063F24	04	195	Hypothetical	112	8×10^{-26}	95/120 (79%)	9497	4DL12-0.71-1.00 C-4AL12-0.43 4BL1-0.71-1.00[a]
43	OSJNBa0063F24	05	759	Purple acid phosphatase	293	1×10^{-79}	400/596 (67%)	9497	4DL12-0.71-1.00 C-4AL12-0.43 4BL1-0.71-1.00[a]
44	OSJNBa0041I05	22	105	Hypothetical	128	6×10^{-31}	87/94 (92%)	7497	5AL12-0.35-0.57
45	OSJNBa0041I05	23	1,146	Transcription activator	120	2×10^{-27}	104/133 (78%)	5726	C-7BL2-0.33
46	OSJNBa0060B06	7	3,294	Hypothetical	302	3×10^{-82}	297/404 (73%)	471	C-4BL1-0.71 C-4DL9-0.31 4A
47	OSJNBb0077M24	6	1,530	Unknown	204	1×10^{-52}	172/211 (81%)	10990	4D 4B 4AL
48	OSJNBb0077M24	7	342	Hypothetical	202	9×10^{-53}	139/151 (92%)	10990	4D 4B 4AL
49	OSJNBa0093J03	22	1,482	Unknown	152	8×10^{-37}	145/185 (78%)	5362	2BL4-0.50-0.89 2DL3-0.49-1.00[a] C-2AL1-0.85
50	OSJNBa0093J03	25	2,424	Lipooxygenase	827	0.0	1,094/1,617 (67%)	10492	4AL13-0.59-0.66 4BS8-0.57-0.81 4DS3-0.67-0.82
51	P0549G05	16	1,479	Beta-1,3 glucanase	302	4×10^{-82}	372/544 (68%)	861	6DL10-0.80-1.00 6AL8-0.90-1.00
52	OSJNBa0072L08	10	1,377	Hypothetical	407	10^{-113}	301/346 (86%)	895	C-4BL1-0.71 C-4DL9-0.31 C-4AL12-0.43
53	OSJNBa0072L08	11	825	Unknown	313	9×10^{-86}	246/280 (87%)	11090	3DL2-0.27-0.81 3AL3-0.42-0.78 3BL2-0.22-0.50 2DL9-0.76-1.00 2BL6-0.89-1.00 C-4AS1-0.20 C-4DS1-0.53 C-2AL1-0.85 4BS4-0.37-0.57
54	OSJNBa0060K21	13	399	Unknown	101	4×10^{-22}	82/101 (81%)	516	4DL13-0.56-1.00[a]
55	P0480H08	8	879	Ribosylation factor	104	1×10^{-22}	110/155 (70%)	11009	C-5BL6-0.29 C-5DL1-0.60 C-5AL12-0.35
56	P0480H08	17	2,298	Unknown	1,106	9×10^{-23}	85/104 (81%)	3812	5AL10-0.57-0.78 5BL9-0.76-0.79 5DL9-0.74-0.76
57	P0480H08	24	819	50S ribosomal protein	277	9×10^{-75}	267/340 (78%)	5520	3DL2-0.27-0.81 3BL10-0.50-0.63

Table 2 (continued)

S. no.	Rice genes information				BLAST search results			Mapped wheat EST	
	BAC/PAC clone	ID	Predicted cDNA size (bp)	Product[a]	Score	*e*-value	% Homology	Contig	Chromosome bin
58	P0704G09	13	4,020	Mla1-like protein	122	3×10^{-27}	315/535 (58%)	14768	6DL10-0.80-1.00 6BL5-0.40-1.00 6AL8-0.90-1.00
59	OSJNBa0074L11	2	4,857	Hypothetical	101	4×10^{-21}	92/121 (76%)	8360	3BL2-0.22-0.50
60	OSJNBa0074L11	9	429	Hypothetical	288	2×10^{-78}	307/423 (72%)	9269	5BL6-0.29-0.75[a] C-5DL1-0.60 5AL12-0.35-0.57
61	OSJNBa0074L11	10	429	Hypothetical	271	4×10^{-73}	295/415 (71%)	9269	5BL6-0.29-0.75[a] C-5DL1-0.60 5AL12-0.35-0.57
62	OSJNBa0074L11	13	1,497	Hypothetical	917	0.0	687/791 (86%)	1308	1BL3-0.85-1.00 1DL2-0.41-1.00
63	OSJNBa0074L11	28	1,608	Phospholipase	473	10^{-133}	600/887 (67%)	10864	4AL5-0.66-0.80 5BL16-0.79-1.00 5A
64	OSJNBa0074L11	31	1,659	Hypothetical	153	3×10^{-37}	275/409 (67%)	1095	C-4DS1-0.53
65	OSJNBa0074L11	35	1,971	Unknown	137	2×10^{-32}	186/276 (67%)	4972	2DS1-0.33-0.47
66	OSJNBa0074L11	39	315	Unknown	115	1×10^{-26}	118/164 (71%)	9485	1DL2-0.41-1.00
67	OSJNBa0074L11	43	447	Pathogenesis-related protein	439	10^{-123}	349/422 (82%)	9485	1DL2-0.41-1.00
68	OSJNBa0085C16	24	1,473	Flavonol glucosyl-transferase	278	5×10^{-75}	251/317 (79%)	3659	C-3BS1-0.33 C-3AS2-0.23 C-3DS3-0.24
69	OSJNBa0085C16	29	2,100	DEAD/DEAH box	110	3×10^{-24}	137/205 (66%)	6174	4AS3-0.76-1.00 4BL5-0.86-1.00
70	OSJNBa0043E10	21	1,872	DEAD/DEAH box	110	3×10^{-24}	137/205 (66%)	6174	4AS3-0.76-1.00 4BL5-0.86-1.00
71	OSJNBa0043E10	26	1,473	UTP glucose transferase	278	5×10^{-75}	251/317 (79%)	3659	C-3BS1-0.33 C-3AS2-0.23 C-3DS3-0.24
72	OSJNBa0034P08	19	1,905	Receptor kinase	500	10^{-141}	473/611 (77%)	8158	C-5DL1-0.60 2BS3-0.84-1.00
73	OSJNBa0034P08	20	1,962	Unknown	163	5×10^{-40}	177/252 (70%)	4624	C-5AL10-0.57[a] C-5DL1-0.60 C-5BL14-0.75[a]
74	OSJNBa0034P08	28	684	Pyruvate decarboxylase	245	3×10^{-65}	227/300 (75%)	11566	4AS1-0.20-0.63 2AS5-0.78-1.00
75	OSJNBa0007P22	5	783	Hypothetical	294	5×10^{-80}	230/270 (85%)	7158	7AS5-0.59-0.89 C-6BS5-0.76 C-4DL9-0.31 7DS4-0.61-1.00 7BS1-0.27-1.00
76	OSJNBa0007P22	43	711	Pyruvate decarboxylase	252	3×10^{-67}	231/304 (75%)	11566	4AS1-0.20-0.63 2AS5-0.78-1.00
77	OSJNBa0042J05	3	1,254	Patatin-like protein	803	0.0	730/944 (77%)	9246	C-2BL2-0.36 C-2DL3-0.49
78	OSJNBa0042J05	7	1,233	Patatin-like protein	478	10^{-135}	464/617 (75%)	9246	C-2BL2-0.36 C-2DL3-0.49

Table 2 (continued)

S. no.	Rice genes information				BLAST search results			Mapped wheat EST	
	BAC/PAC clone	ID	Predicted cDNA size (bp)	Product[a]	Score	*e*-value	% Homology	Contig	Chromosome bin
79	OSJNBa0042J05	14	1,200	Expressed protein AT2g15695.1	190	2×10^{-48}	281/443 (63%)	2382	7DS4-0.61-1.00 7AS5-0.59-0.89 7BS1-0.27-1.00
80	OSJNBa0042J05	24	732	Proteosome alpha subunit	713	0.0	495/541 (91%)	10708	C-4BL1-0.71 C-4DL9-0.31 C-4AL12-0.43
81	OSJNBa0042J05	28	3,696	Unknown	362	10^{-100}	324/410 (79%)	6451	5BL6 5AL10-0.57-0.78
82	OSJNBa0034K24	5	1,081	12.8-kDa protein	101	1×10^{-21}	187/240 (77%)	5361	3BL7-0.63-1.00 3AL5-0.78-1.00 3DL3-0.81-1.00
83	OSJNBa0034K24	6	4,826	26.2-kDa protein	286	7×10^{-77}	359/508 (70%)	7069	7BL10-0.78-1.00 3DS6-0.55-1.00 3A 3BS8-0.78-1.00 7DL2-0.61-0.82
84	OSJNBb0091E08	20	1,020	Cinnamyl-ADH	748	0.0	623/765 (81%)	9239	3BL7-0.63-1.00 3AL5-0.78-1.00 3DL3-0.81-1.00
85	P0038B07	2	1,221	Hypothetical	878	0.0	727/896 (81%)	8841	2BS1-0.53-0.75 C-3AS2-0.23 7AS5-0.59-0.89
86	P0038B07	6	594	Hypothetical	206	2×10^{-53}	199/254 (78%)	6654	C4A C-4DS1-0.53 C-4BS4-0.37
87	OSJNBa0016O23	6	5,049	Helicase	101	5×10^{-21}	92/121 (76%)	8360	3BL2-0.22-0.50
88	OSJNBa0039F06	25	1,539	Cyt. P450	188	9×10^{-48}	601/1,044 (57%)	10064	1BL3-0.85-1.00 1DL2-0.41-1.00 1AL1-0.17-0.61
89	OSJNBa0039F06	29	1,539	Cyt. P450	177	2×10^{-44}	581/1,002 (57%)	4414	1BL3-0.85-1.00 1DL2-0.41-1.00 1AL1-0.17-0.61
90	OSJNBa0095P22	27	1,341	U2SnRNP factor	364	10^{-101}	335/436 (76%)	8474	C-4AL12-0.43 C-4DS1-0.53
91	OSJNBa0095P22	28	1,416	Sugar transporter	301	1×10^{-81}	383/567 (67%)	7227	4BL5-0.86-1.00 4AS3-0.76-1.00 4DL13-0.56-0.71
92	OSJNBa0095P22	29	1,437	Sugar transporter	307	1×10^{-83}	388/573 (67%)	7227	4BL5-0.86-1.00 4AS3-0.76-1.00 4DL13-0.56-0.71
93	OSJNBa0095P22	30	1,431	Sugar transporter	328	8×10^{-90}	392/573 (68%)	7227	4BL5-0.86-1.00 4AS3-0.76-1.00 4DL13-0.56-0.71
94	OSJNBa0095P22	32	1,947	Sugar transporter	302	5×10^{-82}	384/573 (67%)	7227	4BL5-0.86-1.00 4AS3-0.76-1.00 4DL13-0.56-0.71
95	OSJNBa0095P22	36	1,947	Sugar transporter	678	0.0	515/603 (85%)	10646	C-5DS1-0.63 5AS1-0.40-0.75 5BS4-0.43-0.56
96	OSJNBa0095P22	37	2,082	GTP-binding protein	123	4×10^{-28}	94/111 (84%)	10157	4AS1-0.20-0.63

Table 2 (continued)

S. no.	Rice genes information				BLAST search results			Mapped wheat EST	
	BAC/PAC clone	ID	Predicted cDNA size (bp)	Product[a]	Score	*e*-value	% Homology	Contig	Chromosome bin
97	OSJNBa0041L19	6	1,347	Anthranilate-*N*-benzoyltransferase	362	10^{-100}	325/422 (77%)	5763	4AS3-0.76-1.00 4BL5-0.86-1.00
98	OSJNBa0041L19	14	1,338	Anthranilate-*N*-benzoyltransferase	372	10^{-103}	328/422 (77%)	5763	4AS3-0.76-1.00 4BL5-0.86-1.00
99	OSJNBa0090F16	10	1,071	Hypothetical	187	2×10^{-47}	222/327 (67%)	9438	C-4BL1-0.71 4DL13-0.56-0.71
100	OSJNBa0090F16	15	612	Hypothetical	168	5×10^{-42}	192/279 (68%)	4156	C-4BL1-0.71 4DL13-0.56-0.71
101	OSJNBa0090F16	16	1,863	Hypothetical	226	4×10^{-59}	314/486 (64%)	4500	C-4DL9-0.31 4AS1-0.20-0.63
102	P0682E05	32	1,194	Polygalacturonidase	212	5×10^{-55}	246/359 (68%)	6469	5BL16-0.79-1.00 4AL5-0.66-0.80 5DL5-0.76-1.00
103	OSJNBb0039K08	4	1,008	Peroxidase	155	6×10^{-38}	274/446 (61%)	6051	4AS1-0.20-0.63 C-4BL1-0.71 4DL9-0.31-0.56
104	OSJNBb0024E08	23	1,458	Receptor kinase	101	1×10^{-21}	179/225 (60%)	6152	7BL7-0.63-0.78
105	OSJNBa0002C14	25	8,367	Reverse transcriptase	315	3×10^{-85}	293/388 (75%)	6152	7BL7-0.63-0.78
106	OSJNBb0076M06	5	8,346	Reverse transcriptase	315	3×10^{-85}	293/388 (75%)	6152	7BL7-0.63-0.78
107	OSJNBb0076M06	22	807	Hypothetical	190	2×10^{-48}	275/421 (65%)	6152	7BL7-0.63-0.78
108	OSJNBb0076M06	23	603	Hypothetical	477	10^{-135}	392/479 (81%)	6152	7BL7-0.63-0.78
109	OSJNBa0082P17	7	1,632	Unknown	291	9×10^{-79}	401/604 (66%)	7855	6AL8-0.90-1.00 2DS5-0.47-1.00 2BS1-0.53-0.75
110	OSJNBa0082P17	29	1,710	Kinase	278	6×10^{-75}	264/353 (74%)	8158	C-5DL1-0.60 2BS3-0.84-1.00
111	OSJNBa0082P17	32	1,683	U2SnRNP factor	1,285	0.0	916/1,022 (89%)	8474	C-4AL12-0.43 C-4DS1-0.53
112	OSJNBa0064H09	7	1,653	U2SnRNP factor	1,258	0.0	916/1,022 (89%)	8474	C-4AL12-0.43 C-4DS1-0.53
113	OSJNBa0085H07	1	2,871	Disease resistant protein	120	5×10^{-27}	211/337 (62%)	1812	7AS1-0.89-1.00

[a] Besides gene products identifed on the basis of sequence homology, hypothetical and unknown genes are included as defined in Table 1

chromosomes 6 and 8 to group 7 of wheat and rice chromosomes 9 and 12 to group 5 of wheat.

Out of the 113 rice genes showing similarity to bin-mapped wheat ESTs, 97 (85.8%) mapped exclusively on to the three chromosomes of a single wheat chromosome group, suggesting that the distribution of rice genes to more than one wheat homoeologous group predates the evolution of three different wheat genomes. The largest number (35%) of such genes mapped to the wheat chromosomes of group 4, followed by 19% to group 3. The share of genes to wheat chromosomes of groups 1, 2, 5, 6 and 7 was 8%, 7%, 15%, 4% and 12% of the genes, respectively. Approximately 40.7% of the genes were present on all the three homoeologues of a particular wheat chromosome group, whereas others were located on either two or just one of the three homoeologous chromosomes of a group. In fact, out of the 113 genes about one-quarter (27 genes) were present on only 1 of the 21 wheat chromosomes suggesting, at least in some cases, selective elimination of the additional genes in the polyploid wheat genome. Although, it could be explained partly by failure of the chromosome bin location due to a lack of polymorphism for some wheat ESTs (Dr Bikram Gill, personal communication). A mechanism for rapid selective elimination of additional copies of non-coding DNA has been reported in freshly synthesized synthetic wheat polyploids, the so-called process of diploidization (Ozkan et al. 2001). In contrast, 16 of the 113 rice genes were amplified in the wheat genome, with additional copies mapping to chromosomes of more than one homoeologous group. Nine of these 16 genes were present on group 4 wheat chromosomes but with additional copies on other homoeologous groups. Thus, there is clear evidence for the significant movement of genes from

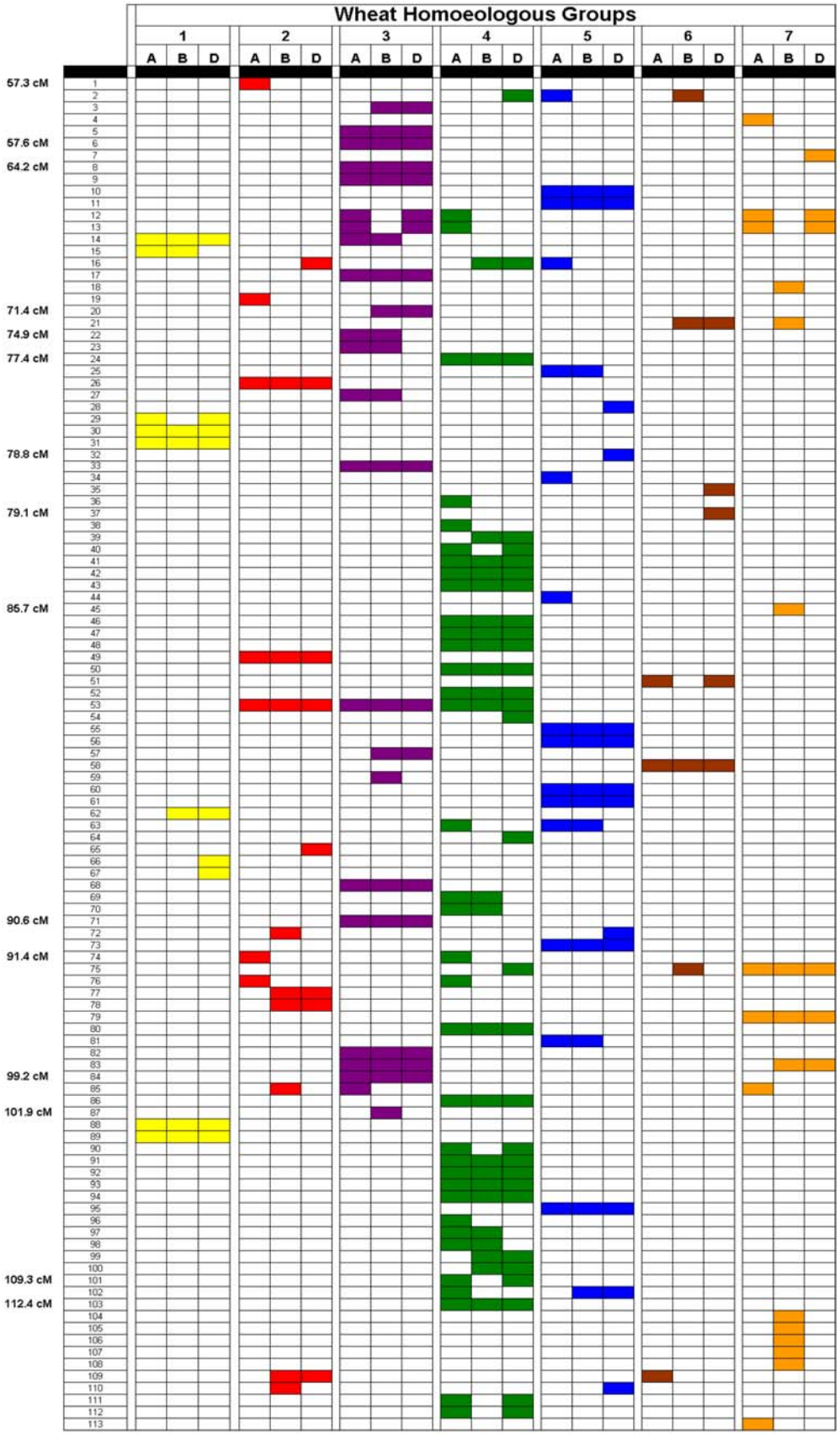


Fig. 1 Depiction of the distribution of rice genes from the long arm of chromosome 11 on all the 21 wheat chromosomes. Serial numbers *on the left* represent the gene order in rice according to Table 2. Each of the seven homoeologous groups of wheat chromosomes has been *color-coded* and matches with rice genes are represented by *color filled cells*. The values in cM represent positions defined by genetically mapped markers present in the BAC/PAC clones used in the analysis (http://rgp.dna.affrc.go.jp/publicdata/estmap2001/Chr11.html

their common ancestral positions to other chromosomal locations. This movement (or duplication onto more than one chromosome group) could have happened either in rice or wheat or partly in both. The genes have moved by both copy-paste (duplication while retaining their ancestral position) and cut-paste (duplication onto another chromosome but elimination from the original position) mechanisms. In several cases, the additional copies of genes resulting from possible segmental duplication or ancient polyploidy have been eliminated (Bowers et al. 2003; Vandepoele et al. 2003). These observations suggest that plant genomes have mechanisms for all the three possibilities, i.e., duplicating, deleting or retaining the original number of genes present in the ancestral species. This is consistent with the previous results that the grass genomes are highly dynamic and that genes have moved from orthologous positions to paralogous positions, where they have been amplified at the same time (Song et al. 2002; Wicker et al. 2003). A series of recent studies have provided evidence for massive genome-wide duplication of genes in both dicots and monocots. It has been postulated that the Angiosperm genomes have a propensity to undergo chromosomal duplication (polyploidization) followed by selective elimination of genes (diploidization) and this has played a major role in the evolution of grass genomes (Paterson et al. 2000; Bowers et al. 2003). In a separate work, Vandepoele et al. (2003) have shown that rice and other cereals are ancient anueploids containing ancient duplications of one or two chromosomes only and not a whole genome polyploid. This duplication has presumably predated the divergence of most cereal crop species as concluded from molecular dating techniques. It is clear from the present analysis that the highest numbers (35%) of rice genes showing significant homology with wheat ESTs are located on the group 4 chromosomes of wheat. Therefore, chromosome 11 of rice and group 4 chromosomes of wheat must share a common evolutionary origin. The remaining two-thirds of the genes have moved away from their ancestral position to different chromosomes by not yet clearly defined mechanisms.

Conservation of gene order between wheat and rice

Though a detailed genetic map of the wheat ESTs to allow a strict comparison of gene orders between rice and wheat is not available, their physical location within specific chromosomal bins does provide a broad view of the order of groups of wheat ESTs along the chromosome. It will be interesting here to analyze whether there is any significant conservation in the order of these genes common between rice chromosome 11 and wheat chromosomes of homoeologous groups 4 and 3, the two most closely related wheat groups. The bin map distribution of 33 rice genes mapping to the wheat chromosomes of homoeologous group 4 and 18 rice genes mapping to wheat homoeologous group 3 is given in Fig. 2a, b. The identity of wheat contigs (contig number) is shown on the left, and the identity of rice genes, including their serial numbers indicative of gene order (in parenthesis) from Table 2, is shown on the right side of the chromosome bar in relation to the corresponding bin position on wheat chromosomes. Genes within a wheat bin could not be ordered, as the exact position of ESTs within a bin is not yet established; however, we have tried to arrange them within a bin manually in such a way as to maximize the conservation of gene order between rice and wheat. Despite this, the comparison of gene order in rice with the order of mapped ESTs in the wheat chromosome bins of group 4 and 3 shows a high degree of rearrangements and, apparently, there is very little conservation of gene order or colinearity. It would be desirable to experimentally establish the order of genes within a wheat bin, as suggested earlier (Han et al. 1999; Bennetzen and Ramakrishna 2002). Nonetheless, it seems from the present analysis that genes have shifted not only to different chromosomes from their ancestral position but also to different positions on the same chromosomes, thus affecting their gene order. For example, out of the total 29 wheat contigs (33 rice genes) mapping on the group 4 chromosomes, only 2 could be aligned with respect to their positions on 4A, 4B and 4D. Three wheat contigs, numbers 9887, 8719 and 7298, located at the end of chromosome 4A short arm (top of Fig. 2a) have shifted to near the end of the long arm of chromosome 4B and are completely missing from chromosome 4D (bottom of Fig. 2a). These discrepancies between the location of genes on the chromosomes of A, B and D genomes of wheat must have occurred after the divergence of wheat from the common progenitor of rice and wheat. The genome of hexaploid wheat is reported to have a number of such translocations especially in homoeologous groups 4, 5 and 7 (Sorrells et al. 2003). The situation was better with group 3 chromosomes where all the eight contigs located on 3A, 3B and 3D could be aligned with each other. However, there was no significant conservation of gene order between rice and wheat, as the serial number of rice genes on the right hand side of the chromosome bar do not follow any definite ascending or descending order in either group 4 or group 3 (Fig. 2a, b). This is in sharp contrast to a very high level of conservation of gene order between chromosome 1 of rice and chromosome 3 of sorghum, another important member of the grass family (Klein et al. 2003). It also points to the possibility that the genomes of rice, sorghum and wheat have evolved at different rates with respect to the rearrangement of gene order.

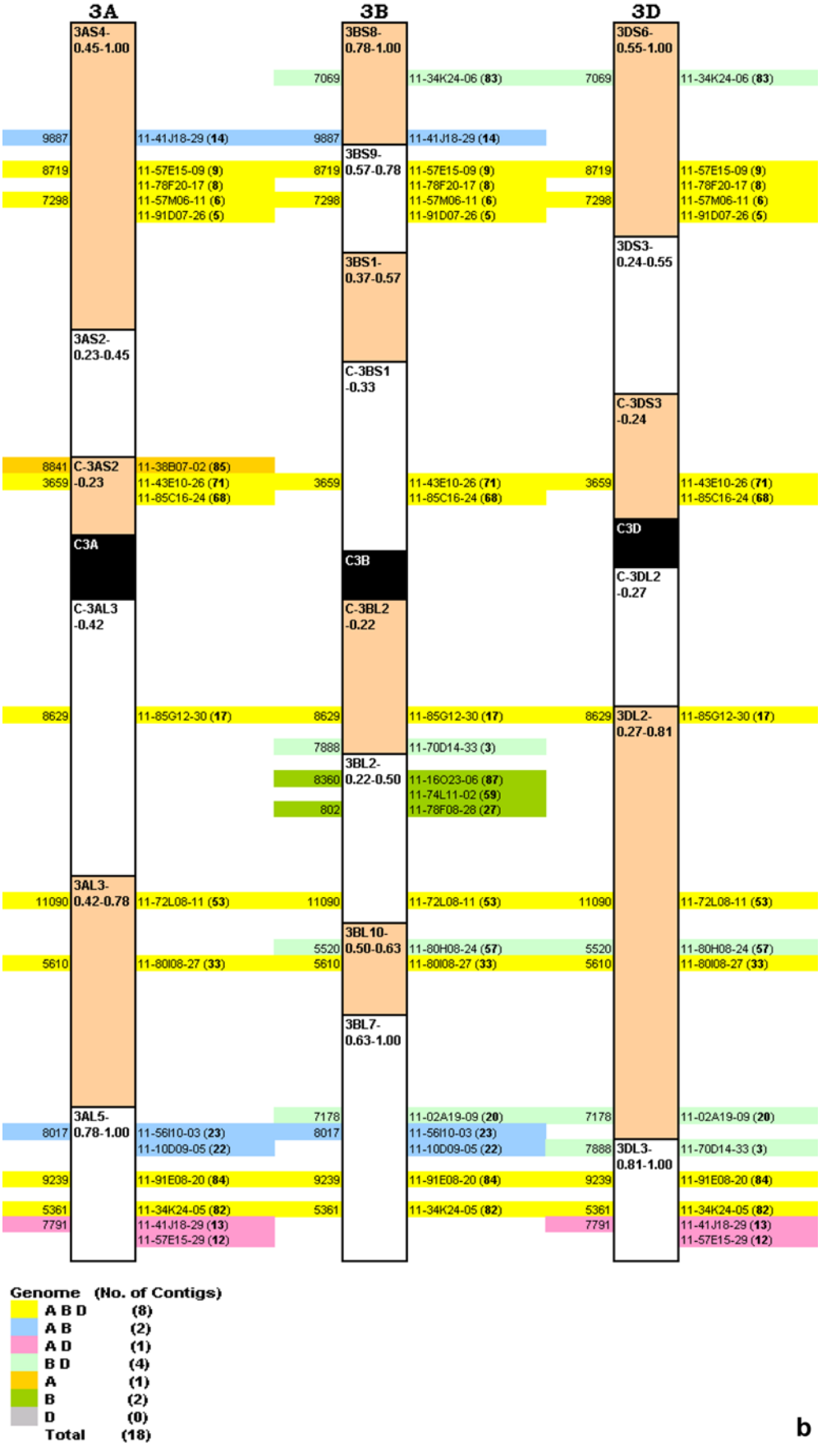
The pattern of gene order in the rice-wheat genome comparison presented here differs from macro-level colinearity reported earlier by several workers based on

Fig. 2 Location of genes from the long arm of rice chromosome 11 (57.3–116.2 cM) on group 4 (**a**) and group 3 (**b**) wheat chromosomes based on matches with bin-mapped wheat ESTs at a cutoff bit score of 100 and *e* value of 10^{-20}. The identity of wheat contigs is shown *on the left* and that of the rice gene *on the right* side of the chromosome bar. The identity of rice genes given on the right side includes chromosome number, BAC/PAC name, gene ID and serial number (gene order) from Table 2 ▶


4A
4B
4D
4AS3-0.76-1.00
4AS4-0.63-0.76
4AS1-0.20-0.63
C-4AS1-0.20
C4A
C-4AL12-0.43
4AL12-0.43-0.59
4AL13-0.59-0.66
4AL5-0.66-0.80
4AL4-0.80-1.00
4BS1-0.81-1.00
4BS8-0.57-0.81
4BS4-0.37-0.57
C-4BS4-0.37
C4B
C-4BL1-0.71
4BL1-0.71-0.85
4BL5-0.86-1.00
4DS2-0.82-1.00
4DS3-0.67-0.82
4DS1-0.53-0.67
C-4DS1-0.53
C4D
C-4DL9-0.31
4DL9-0.31-0.56
4DL13-0.56-0.71
4DL12-0.71-1.00
11-95P22-28-32 (91-94)
11-41L19-06,14 (90,97)
11-43E10-21 (70)
11-85C16-29 (69)
11-05H02-10 (40)
11-95P22-37 (96)
11-07P22-43 (76)
11-34P08-28 (74)
11-39K08-04 (103)
11-90F16-16 (101)
11-72L08-11 (53)
11-38B07-06 (86)
11-64H09-04 (112)
11-82P17-32 (111)
11-95P22-27 (90)
11-42J05-24 (80)
11-63F24-4,5 (42,43)
11-63F24-03 (41)
11-07D07-02 (38)
11-41J17-24 (36)
11-56I10-23 (24)
11-72L08-10 (52)
11-82E05-32 (102)
11-74L11-28 (63)
11-41J18-01 (13)
11-57E15-29 (12)
11-93J03-25 (50)
11-60B06-07 (46)
11-90F16-15 (100)
11-90F16-10 (99)
11-05H02-09 (39)
11-85G12-27 (16)
11-74L11-31 (64)
11-64H09-04 (112)
11-82P17-32 (111)
11-95P22-27 (90)
11-41C22-24 (2)
Genome (No. of Contigs)
A B D (10)
A B (2)
A D (3)
B D (5)
A (6)
B (0)
D (3)
Total (29)

a

Fig. 2 (continued)



a limited number of RFLP marker sets (Gale and Devos 1998) and favors conclusions drawn by Gallego et al. (1998), Gaut (2002) and Sorrells et al. (2003) that there is very little conservation of gene order between rice and wheat. A large number of homologous wheat ESTs have not been bin-mapped as yet but these can also be expected to follow the same pattern of organization as the 113 bin-mapped ESTs reported here. Since the genetic order of ESTs within a wheat bin is not known, this makes the comparison less stringent. Once the detailed genetic order of wheat ESTs becomes available this will pose a further restriction on matching the order of genes between rice and wheat. However, our results do not rule out the presence of large chromosomal segments with conserved gene order between wheat and other rice chromosomes. Also it does not rule out the presence of smaller orthologous segments, which will become apparent only after getting the high-density genetic map of wheat. A clearer picture will emerge with an increasing number of bin-mapped wheat ESTs becoming available.

To conclude, the comparative analysis of genes predicted from the genomic sequence of the long arm of rice chromosome 11 with the wheat EST database shows significant conservation of individual gene sequences between rice and wheat. Furthermore, it is shown that rice chromosome 11 and wheat chromosomes of homoeologous group 4 share a common origin. This complements the earlier work of Sorrells et al. (2003) where wheat chromosomes related to each of the 12 rice chromosomes, except 9, 11 and 12, were clearly identified. However, the order of rice genes in relation to distribution of their homologues in the wheat genome emerges as a complex pattern because location of these genes seems to have shifted between and within the chromosomes. Nearly two-thirds of the rice genes from chromosome 11 are distributed on the remaining six homoeologous groups of wheat indicating a high level of rearrangements in the grass genomes. It would be useful to further evaluate the implications of the high level of complexity for synteny analysis of wheat and rice genomes, both at macro and micro-levels. Such findings are important for developing molecular breeding strategies and for understanding the genome evolution in cereals using the rice genome as a model.

Acknowledgements The support of the Department of Biotechnology, Government of India, and the Indian Council of Agricultural Research is acknowledged. We are thankful to NCBI, USDA and RGP for making their databases and genome analysis tools publicly available as well as for clarifications of our queries by e-mails. Technical assistance of G. Chitra, S. Sahoo, V. Gupta, M. Francis and A. Sankar is appreciated.

References

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

Appels R, Francki M, Chibbar R (2003) Advances in cereal functional genomics. Funct Integr Genomics 3:1–24

Bennetzen JL, Freeling M (1993) Grasses as a single genetic system: genome composition, colinearity and compatibility. Trends Genet 9:259–261

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

Bowers JE, Chapman BE, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433–537

Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S, Higingbottom S, Phimphilai J, Phimphilai D, Thurmond S, Gaudette B, Li P, Liu J, Hatfield J, Main D, Farrar K, Henderson C, Barnett L, Costa R, Williams B, Walser S, Atkins M, Hall C, Budiman MA, Tomkins JP, Luo M, Bancroft I, Salse J, Regad F, Mohapatra T, Singh NK, Tyagi AK, Soderlund C, Dean RA, Wing RA (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14:537–545

Deynze AE Van, Dubcovsky J, Gill KS, 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

Faris JD, Friebe B, Gill BS (2002) Wheat genomics: exploring the polyploid model. Curr Genomics 3:577–591

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95:1071–1974

Gallego F, Feuillet C, Messmer M, Penger A, Graner A, Yano M, Sasaki T, Keller B (1998) Comparative mapping of the two wheat leaf rust resistance loci *Lr1* and *Lr10* in rice and barley. Genome 41:328–336

Gaut BS (2002) Evolutionary dynamics of grass genomes. New Phytol 154:15–28

Han F, Kilian A, Chao JP, Kudrna D, Steffenson B, Yamamoto K, Matsumoto T, Sasaki T, Kleinhofs A (1999) Sequence analysis of a rice BAC covering the syntenous barley Rpg1 region. Genome 42:1071–1076

Heslop-Harrison JS (2000) Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. Plant Cell 12:617–635

Islam AKMR, Shepherd KW (1992) Substituting ability of individual barley chromosomes for wheat chromosomes 1 substitutions involving barley chromosomes 1, 3 and 6. Plant Breed 109:141–150

Kihara H (1924) Cytologische und genetische Studien bei wichtigen Getreidearten mit besonderer Ruckischt auf das Verhalten der Chromosomen und die Sterilitat in den Bastarden. Mem Coll Sci Kyoto Imp Univ Ser B 1:1–200

Klein PE, Klein RR, Vrebalov J, Mullet JE (2003) Sequence-based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement. Plant J 34:605–621

Kurata N, Moore G, Nagamura Y, Foote T, Yano M, Minobe Y, Gale M (1994) Conservation of genome structure between rice and wheat. Bio/Technology 12:276–278

Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. Plant Cell 13:1735–1747

Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang C-X, Katsar CS, Lan TH, Lin YR, Ming R, Wright RJ (2000) Comparative genomics of plant chromosomes. Plant Cell 12:1523–1539

Qi L, Echalier B, Friebe B, Gill BS (2003) Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs. Funct Integr Genomics 3:39–55

Sakata K, Nagamura Y, Numa H, Antonio BA, Nagasaki H, Idonuma A, Watanabe W, Shimizu Y, Horiuchi I, Matsumoto T, Sasaki T, Higo K (2002) RiceGAAS: an automated annotation system and database for rice genome sequence. Nucleic Acids Res 30:98–102

Sears ER (1941) Chromosome pairing and fertility in hybrids and amphidiploids in the Triticinae. Mo Agric Exp Stn Res Bull 337:1–20

Song R, Llaca V, Messing J (2002) Mosaic organization of orthologous sequences in grass genomes. Genome Res 12:1549–1555

Sorrells ME, Rota ML, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Mahmoud MA, 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 SF, Choi D-W, 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

Vandepoele K, Simillion C, Van de Peer Y (2003) Evidence that rice and other cereals are ancient aneuploids. Plant Cell 15:2192–2202

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu Z-D, Dubcovsky J, Keller B (2003) Rapid genome divergence at orthologous low molecular weight glutelin loci of the A and A^M genomes of wheat. Plant Cell 15:1186–1197