

Dispersion and domestication shaped the genome of bread wheat

Paul J. Berkman^{1,2,3}, Paul Visendi^{1,2}, Hong C. Lee¹, Jiri Stiller^{1,3}, Sahana Manoli¹, Michał T. Lorenc^{1,2}, Kaitao Lai^{1,2}, Jacqueline Batley¹, Delphine Fleury⁴, Hana Šimková⁵, Marie Kubaláková⁶, Song Weining⁶, Jaroslav Doležel⁵ and David Edwards^{1,2,*}

¹School of Agriculture and Food Sciences, University of Queensland, Brisbane, QLD, Australia

²Australian Centre for Plant Functional Genomics, University of Queensland, Brisbane, QLD, Australia

³CSIRO Plant Industry, St Lucia, QLD, Australia

⁴Australian Centre for Plant Functional Genomics, University of Adelaide, Glen Osmond, SA, Australia

⁵Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Olomouc, Czech, Republic

⁶State Key Laboratory of Crop Stress Biology in Arid Areas, College of Agronomy and Yangling Branch of China Wheat Improvement Center, Northwest A&F University, Yangling, Shaanxi, China

Received 10 October 2012;

revised 10 December 2012;

accepted 11 December 2012.

*Correspondence (Tel +61 0 7 3346 7084;

fax +61 0 7 3365 1176; email

Dave.Edwards@uq.edu.au)

Summary

Despite the international significance of wheat, its large and complex genome hinders genome sequencing efforts. To assess the impact of selection on this genome, we have assembled genomic regions representing genes for chromosomes 7A, 7B and 7D. We demonstrate that the dispersion of wheat to new environments has shaped the modern wheat genome. Most genes are conserved between the three homoeologous chromosomes. We found differential gene loss that supports current theories on the evolution of wheat, with greater loss observed in the A and B genomes compared with the D. Analysis of intervarietal polymorphisms identified fewer polymorphisms in the D genome, supporting the hypothesis of early gene flow between the tetraploid and hexaploid. The enrichment for genes on the D genome that confer environmental adaptation may be associated with dispersion following wheat domestication. Our results demonstrate the value of applying next-generation sequencing technologies to assemble gene-rich regions of complex genomes and investigate polyploid genome evolution. We anticipate the genome-wide application of this reduced-complexity syntenic assembly approach will accelerate crop improvement efforts not only in wheat, but also in other polyploid crops of significance.

Keywords: *Triticum aestivum*, genome sequencing, evolution.

Introduction

Wheat is a major food crop and is used widely for making breads, pastries, noodles and dumplings. It was domesticated around 10 000 years ago in both tetraploid and hexaploid forms (Dubcovsky and Dvorak, 2007), and today, hexaploid bread wheat (*Triticum aestivum*) provides roughly a fifth of world's food (Food and Agriculture Organisation of the United Nations, 2012). Genome analysis in bread wheat is a challenge as it has a genome nearly six times larger than the human genome and consists of between 80 and 90% repetitive sequence (Wanjugi *et al.*, 2009). Bread wheat is also hexaploid with 21 pairs of chromosomes, being derived from a combination of three diploid donor species each with seven pairs of chromosomes. The donor species are proposed to have diverged from an ancestral diploid species between 2.5 and 6 MYA (Chantret *et al.*, 2005; Huang *et al.*, 2002). Hexaploid bread wheat evolved through two interspecific hybridization events, each accompanied by polyploidization. The first, occurring between 0.5 and 3 MYA, combined the genomes of *Triticum urartu* (AuAu) and an unidentified species (BB) highly similar to *Aegilops speltoides* to produce the allotetraploid genome of wild emmer wheat or *Triticum turgidum* (AuAuBB)

(Chantret *et al.*, 2005; Eckardt, 2001; Huang *et al.*, 2002). The second event combined the genomes of *T. turgidum* (AuAuBB) and *Aegilops tauschii* (DD) to produce the allohexaploid genome of *T. aestivum* (AuAuBBDD) (McFadden and Sears, 1946). Each diploid progenitor genome is around 5500 million base pairs, almost twice the size of the human genome, and consists of between 80% and 90% repetitive elements (Dvořák, 2009). A greater number of genes for domestication traits are found on the A and B genomes (Gegas *et al.*, 2010), suggesting that the tetraploid was domesticated prior to the emergence of the hexaploid. No wild hexaploid wheats are known, and it is accepted that *T. aestivum* originated from a cross between *Ae. tauschii* and domesticated tetraploid emmer, probably South or West of the Caspian Sea around 8000 years ago (Giles and Brown, 2006; Nesbitt and Samuel, 1995; Salamini *et al.*, 2002).

The evolution of the homoeologous gene space in polyploid genomes can be investigated by comparative genomics approaches, including differential gene loss and single-nucleotide polymorphisms (Bekaert *et al.*, 2011; Thomas *et al.*, 2006). In addition, the group 7 chromosomes are known to contain QTL associated with boron tolerance, drought tolerance and pathogen resistance (Dolores Vazquez *et al.*, 2012; Genc *et al.*, 2010;

Please cite this article as: Berkman, P.J., Visendi, P., Lee, H.C., Stiller, J., Manoli, S., Lorenc, M.T., Lai, K., Batley, J., Fleury, D., Šimková, H., Kubaláková, M., Weining, S., Doležel, J. and Edwards, D. (2013) Dispersion and domestication shaped the genome of bread wheat. *Plant Biotechnol. J.* doi: 10.1111/pbi.12044

Schnurbusch *et al.*, 2007). Consequently, while the group 7 has been sequenced as study of genome evolution, sequencing this chromosome group also represents an opportunity to pursue agriculturally important traits.

Results

Using second-generation sequencing (2GS) of isolated group 7 chromosome arms, as per our previously described methodology (Berkman *et al.*, 2011, 2012), we have assembled genomic regions containing all or nearly all genes for these chromosomes and ordered and aligned the majority of these genes based on synteny with *B. distachyon*, *O. sativa*, and *S. bicolor* (Table 1, Table S1). We identified 9258 genes in total to be present on the three chromosomes, 5532 (59.8%) of which were placed into the syntenic builds (Table S1).

We obtained from the GrainGenes database cDNA sequences representing 18 785 loci (Carollo *et al.*, 2005; Matthews *et al.*, 2003) that were located to defined chromosomal regions by hybridization to DNA from wheat deletion lines (Hossain *et al.*, 2004). These cDNAs were reciprocal best BLAST aligned with each of the group 7 assemblies (Table S4). The results are consistent with our previous results for 7DS and the expected error rate for bin mapping cDNAs in wheat (Berkman *et al.*, 2011). The results showed group 7 mapped ESTs preferentially aligning with the group 7 chromosome assemblies. Some sequences within the 7A and 7D assemblies are also aligned with 4A, which is consistent with the previously characterized ancestral translocation between 7BS and 4AL (Berkman *et al.*, 2012; Hernandez *et al.*, 2012) (Figure S2). Genetic markers from *Ae. tauschii* chromosome 7 (Luo *et al.*, 2009) demonstrated a high degree of concordance with the order and orientation of the 7D syntenic build (Figure S3), as per our previous results (Berkman *et al.*, 2011).

Comparison of assemblies demonstrated that the majority of genes remain conserved, with a copy present on each of the 3 homoeologous chromosomes (Figure 1). Within the syntenic builds, we identified 1291 genes to be present on all three chromosomes and were orthologous to predicted syntenic *B. distachyon* genes. In addition, we identified 550 *B. distachyon* genes with orthologs present on only 2 wheat genomes and 545 *B. distachyon* genes with orthologs present on only a single wheat genome. Greater gene loss was observed in the A and B genomes compared with the D genome, with 2988, 2905 and 3351 genes identified in the A, B and D genomes, respectively (Figure 1, Table S2).

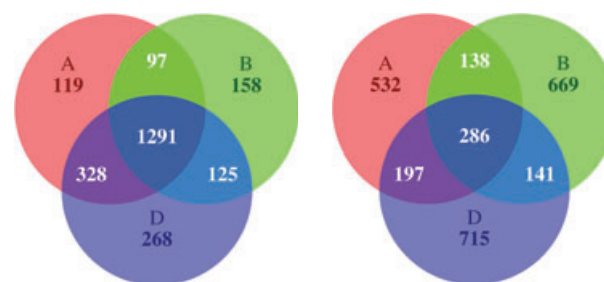


Figure 1 Venn diagram representing genes present on 7A (red), 7B (green) and 7D (blue). Genes within the syntenic build are represented on the left. Genes not found within syntenic regions of related grass species are represented on the right.

We predicted the gene ontology (GO) terms for genes retained in the wheat group 7 chromosomes compared with all genes present in syntenic regions of *O. sativa*, *S. bicolor* and *B. distachyon*. GO terms with a STRING (Szklarczyk *et al.*, 2011) enrichment score >1 are listed in Table 2. The genes retained in the three subgenomes present differing GO profiles in preferentially retained genes when ranked by enrichment index (Table 2, Table S3). The A and B subgenomes have preferentially retained genes relating to basic cellular processes, while the D subgenome has maintained a greater proportion of genes related to specialized processes.

Protein interaction network analysis was undertaken based on the protein interaction network of *Arabidopsis thaliana*. This revealed an abundance of genes from both 7A and 7B involved in gene networks of ribosomal proteins, transcription factors and RNA polymerases and a reduced representation of these large networks observed for 7D (Figure S1). Network analysis also revealed an increased proportion of 7D genes not associated with a network, as well as the presence of a proteolysis gene network not observed in 7A or 7B (Table 2, Figure S1).

By comparing whole-genome sequence data for four Australian wheat varieties with the group 7 assemblies, we predicted more than 900 000 intervarietal SNPs on the assembled group 7 contigs with an accuracy of 93% (Lorenz *et al.*, 2012). Syntenic builds for chromosomes 7A and 7B showed significantly greater intervarietal SNP polymorphism than 7D (Figure 2, Table 3), with 14 059, 9396 and 3137 SNPs identified on syntenic builds 7A, 7B and 7D, respectively.

Chromosome arm	Velvet assembly statistics			Syntenic build statistics			
	Size (Mbp) [†]	2GS data coverage (×)	N50	Total (Mbp)	Longest (bp)	Length (Mbp)	Gene density (per Mbp on chromosome)
7AS	407	27.96	2503	210.57	49,550	6.85	3.55
7AL	407	46.04	2893	256.46	51,272	7.75	3.80
7BS	360	44.81	3596	220.93	56,148	6.62	3.29
7BL	540	27.96	1929	253.91	36,580	5.97	3.19
7DS	381	72.62	2098	210.62	49,276	7.47	4.45
7DL	346	76.68	5194	238.07	59,194	13.48	4.79

[†]Safāi *et al.* (2010).

Table 1 Syntenic build summary table

Table 2 Enriched GO terms of subgenomes against hexaploid gene content

7A		7B		7A & 7B (tetraploid)		7D	
Enrichment score	GO Term	Enrichment score	GO Term	Enrichment score	GO Term	Enrichment score	GO Term
2.72	Primary metabolism	1.64	Cytoplasmic protein	1.84	Cytoplasmic protein	1.44	DNA replication and transposition
1.8	Cytoplasmic protein	1.28	DNA metabolism	1.36	Primary metabolism	1.29	Reproductive structure development
1.47	Chloroplast protein	1.09	Nucleolar lumen protein	1.18	Chloroplast envelope protein	1.26	Metal ion binding
1.18	Proteolysis						
1.06	DNA metabolism						

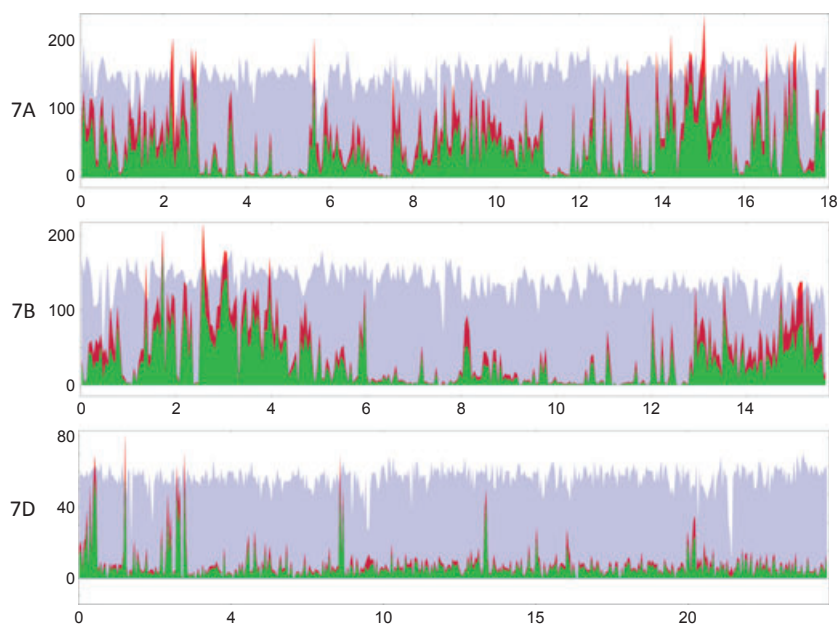


Figure 2 Single-nucleotide polymorphism (SNP) distribution across the syntenic builds of 7A, 7B and 7D. Raw SNP density numbers (SNPs/50 kbp) are represented in green, SNP densities normalized for coverage are represented in red, and read coverage is represented in blue. The X-axis represents the position on the syntenic build (Mbp) from the short-arm telomere (left) to the long-arm telomere (right).

Discussion

Wheat has a complex evolutionary history, with the tetraploid forming in the wild, and later combining with the D genome species to produce hexaploid bread wheat. By sequencing and assembling the low copy and unique gene regions of isolated chromosome arms representing homoeologous chromosomes, we can observe the impact of this history on genome structure.

Gene loss is common following whole-genome duplication (WGD) (Paterson *et al.*, 2009; Schnable *et al.*, 2009; Thomas *et al.*, 2006; Woodhouse *et al.*, 2010). WGD also provides the opportunity for the neo- or subfunctionalization of duplicated genes, and this has been suggested as an explanation of the prevalence of polyploidy in plants (Freeling and Thomas, 2006). In *Brassica rapa*, ancient WGD events have been identified as the basis for the increased morphological plasticity observed in the *Brassica* family (Wang *et al.*, 2011). The dosage balance hypothesis suggests that gene networks in plants significantly influence the process of gene loss following genome duplication (Thomas *et al.*, 2006). While the dosage balance hypothesis holds true in a constant selective environment, recent evidence suggests that a change in the selective environment at the time of WGD can result in the emergence of new genes or functions that provide a specific selective advantage (Bekaert *et al.*, 2011). By using the same method to generate each of the syntenic builds, the

observed differential retention of genes between chromosomes represents true differences in gene content between the genomes.

Greater gene loss was observed in the A and B genomes compared with the D genome, (Figure 1, Table S2) which, when combined with the smaller chromosome size of 7D, appears to have resulted in a higher overall gene density on 7D (Table 1). This reflects two rounds of fractionation. Initial gene loss occurred in the A and B genomes following tetraploidy between 0.5 and 3 MYA, with subsequent loss from all three genomes following the formation of the hexaploid around 8000 years ago. The genes retained in the three subgenomes present differing GO profiles in preferentially retained genes when ranked by enrichment index (Table 2, Table S3). This suggests differential selection pressure between the first and second polyploidization events. Chromosome 7D possesses enriched GO terms related to metal ion

Table 3 Subgenomic varietal SNP profiles from four Australian cultivars

Chromosome	# SNPs	SNPs/Mb
7A	14059	963.3
7B	9396	746.2
7D	3137	149.7

processing and reproductive development, which is consistent with a recent suggestion of the D genome's contribution to adaptability and dispersion of hexaploid bread wheat (Dubcovsky and Dvorak, 2007).

Changes in the selective environment may cause the retention of genes that provide a selective advantage, with the absolute dosage of advantageous genes taking precedence over the dosage balance of highly networked genes which applies under 'normal' conditions (Bekaert *et al.*, 2011). Protein interaction network analysis revealed a reduced representation of large networks for 7D and an increased proportion of 7D genes not associated with a network (Table 2, Figure S1). Following from the above-described work of Bekaert *et al.* (2011), the differences we observed in the gene network profile of the D genome, in which minimally networked genes are preserved, suggest changes in the selective pressure of wheat following hexaploidization. This is consistent with the shift from natural to human selection following *T. turgidum*/*A. tauschii* hybridization associated with domestication and the expansion of environments in which hexaploid wheat was grown. The differential gene loss between the A, B and D genomes further supports the recently proposed theories for selective gene retention and loss following polyploidization (Bekaert *et al.*, 2011). Our results therefore suggest that following the emergence of *T. aestivum*, the expansion of hexaploid wheat to new environments has driven genome fractionation and evolution in this species.

The abundance and type of single-nucleotide polymorphisms (SNPs) in a genome reflect the history of selection and evolution in the species. By comparing whole-genome sequence data for four Australian wheat varieties with the group 7 assemblies, we identified more than 900,000 SNPs on the assembled group 7 contigs. Syntenic builds for chromosomes 7A and 7B showed significantly greater intervarietal SNP polymorphism than 7D (Figure 2, Table 3). This is consistent with previous results (Chao *et al.*, 2009). Following the origin of *T. aestivum*, the tetraploid and hexaploid existed in sympatry for a period (Dvorak *et al.*, 1998), and it has been suggested that gene flow between the two species resulted in greater sequence diversity within the A and B subgenomes when compared with the D subgenome (Caldwell *et al.*, 2004; Dvorak *et al.*, 2006; Talbert *et al.*, 1998). The greater genetic diversity we observe on the A and B genomes is most likely due to early gene flow occurring between hexaploid *T. aestivum* (AuAuBBDD) and its tetraploid progenitor *T. turgidum* (AuAuBB), without a similar flow occurring between the hexaploid and *Ae. tauschii* (DD) (Dvorak *et al.*, 2006).

Summary

The greater levels of gene conservation, together with the relative abundance of non-networked genes, suggests that the D genome has played an important role in bread wheat evolution and that the hexaploid wheat genome has been shaped by domestication and dispersal. This analysis also demonstrates the utility of second-generation sequencing technologies in the analysis of complex polyploid genomes without a reference genome sequence. While the absence of a reference genome sequence hampers investigation of issues of biological and economic importance in polyploid crops, this study demonstrates that new sequencing technologies can yield important insight into polyploid genome evolution and crop physiology, prior to the availability of a complete genome sequence.

Experimental procedures

Data generation, assembly and validation

Professor Bikram Gill (Kansas State University, Manhattan, USA) provided seeds of double ditelosomic lines for the group 7 chromosomes of *Triticum aestivum* cv. Chinese Spring. The seeds were germinated and root tips of young seedlings were used for the preparation of liquid suspensions of intact chromosomes as previously described (Vrána *et al.*, 2000). Group 7 chromosome arms were flow-sorted as telocentric chromosome in batches of 19–39 000 chromosomes representing 20–30 ng DNA. To estimate contamination with other chromosomes, 1000 chromosomes were sorted onto a microscope slide in three replicates and used for fluorescence *in situ* hybridization (FISH) with probes for *Afa* family and telomeric repeats (Kubaláková *et al.*, 2005). Purity in sorted fractions varied from 84% (7AS, 7AL, 7DS) to 94% (7BS, 7BL, 7DL). Chromosomal DNA was purified and subsequently amplified using the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Chalfont St. Giles, UK). A total of 200 ng of pooled, amplified DNA from each chromosome arm was used to prepare Illumina paired-end libraries, which were sequenced on the Illumina GAlIx and Illumina HiSeq2000 platforms as previously described (Berkman *et al.*, 2011, 2012). All *B. distachyon*, *O. sativa* and *S. bicolor* gene sequences excluding intron regions were downloaded from Phytozome version 7 (Goodstein *et al.*, 2012).

Details of the paired-end data generated on Illumina GAlIx and HiSeq 2000 for each chromosome arm is presented in Table S1. All data have been submitted to the NCBI short-read archive, references SRA025100.1, SRA028115.1, SRA049415.1, SRA049416.1, SRA049417.1 and SRA049418.1. All sequence data were filtered and trimmed using an in-house script, trimConverter.py, to produce reads with a Phred quality score of at least 20 at each nucleotide position and a minimum length of 63 bp. The trimmed and filtered read set was assembled using Velvet version 1.1.04 (Zerbino and Birney, 2008) on a DELL R905 server with 128 GB RAM. Each Velvet assembly used a kmer size of 63 bp and an expected coverage representing the read depth after filtering (Table S1). The assembly statistics and predicted size of each chromosome arm are displayed in Table S1 with more detailed assembly statistics presented in Table S1.

Producing syntenic builds and assembly annotation

A comparative genomics approach was applied to order and orientate the wheat contigs into a draft syntenic build, as previously described (Berkman *et al.*, 2011, 2012; Mayer *et al.*, 2009), with results presented in Table S2A. In this approach, *B. distachyon*, *O. sativa* and *S. bicolor* were used as the genomes for syntenic comparison, with this order of evolutionary closeness to wheat determining the decision hierarchy in the algorithm. In the case of infrequent tandem gene duplications, the duplications were assembled individually and placed adjacent to each other in the final syntenic build. The process of producing the syntenic build also identified wheat homologs to *B. distachyon* genes outside of the predicted syntenic regions and reflect previously described gene movement (Berkman *et al.*, 2012; Wicker *et al.*, 2010, 2011) (Table S2B). Syntenic builds for each chromosome arm were combined to form a single syntenic build for each chromosome, with the remaining assembled contigs placed into a chromosome-specific 'extra_contigs' files. The full syntenic chromosome sequences of *B. distachyon* version 192, *O. sativa*

version 193 and *S. bicolor* version 79 were all downloaded from Phytozome (Goodstein *et al.*, 2012; International Rice Genome Sequencing Project, 2005; Paterson *et al.*, 2009; Vogel *et al.*, 2010). Following the generation of the syntenic builds, the number of genes on each of the group 7 chromosomes was calculated as per the methodology previously described (Berkman *et al.*, 2011, 2012). The majority of contigs from each short-read assembly did not match any *B. distachyon* gene, and reviewing the annotation of these contigs suggests that they are predominantly made up of nested transposable element insertions (data not shown), consistent with our previous results (Berkman *et al.*, 2011, 2012).

We annotated each syntenic build with predicted genes, gene functional annotations, as well as homoeologous and varietal SNPs. Predicted genes were identified by comparing all *B. distachyon* genes with the syntenic builds using WU-BLASTN (<http://hg.wustl.edu/info/README.html>) with an E-value cut-off of 1e-5 and a minimum distance between HSPs of 4 kbp. Gene functional annotations were predicted by comparing each syntenic build and its extra contigs with the Uniref90 database using BLASTX (Altschul *et al.*, 1990) with an E-value cut-off of 1e-10. Homoeologous SNPs were detected by comparison of chromosome arm syntenic builds to raw short-read sequences using custom in-house software SGSautoSNP, while varietal SNPs were detected by aligning varietal WGS data from four Australian cultivars (Gladius, Drysdale, Excalibur, and RAC875) using the same tool. All annotations and sequences are available for public access at www.wheatgenome.info (Lai *et al.*, 2012).

Comparison of homoeolog gene content

Wheat gene sequences predicted by comparison of the chromosome assemblies with *B. distachyon* genes were compared with the Swissprot database (downloaded 16th August 2011) using BLASTX with an E-value cut-off of 1e-5. The same comparison with Swissprot was performed with all predicted genes from *B. distachyon* genome as well as a subset of *B. distachyon* genes within regions syntenic with wheat chromosome 7. The resulting lists of UNIPROT accession numbers were analysed using the DAVID GO functional annotation clustering tool (Huang *et al.*, 2008, 2009) with the background of the complete *B. distachyon* genome as a control measure of enrichment for this particular genomic region in *B. distachyon*. UNIPROT accession numbers identified from 7A, 7B and 7D, as well as a combined list of all three were compared, with the background set of all genes from the *B. distachyon* syntenic region. Annotations of COG_ONTOLOGY in the Functional_Categories database as well as GOTERM_BP_ALL, GOTERM_BP_FAT, GOTERM_CC_ALL, GOTERM_CC_FAT, GOTERM_MF_ALL, and GOTERM_MF_FAT in the Gene_Ontology database were considered in characterizing enriched function.

Brassica distachyon gene sequences with an ortholog on each of the 7A, 7B and 7D syntenic builds were identified as per our previously described methodology (Berkman *et al.*, 2011, 2012). The full set of *Arabidopsis thaliana* coding regions was downloaded from Phytozome version 7 (Goodstein *et al.*, 2012) and compared with the *B. distachyon* gene sequences using BLASTN (Altschul *et al.*, 1990) with an E-value cut-off of 10e-05. The list of names for the *A. thaliana* genes with top hits against the three sets of *B. distachyon* genes was then loaded into STRING (Szklarczyk *et al.*, 2011) to identify the representation of networked genes on each group 7 chromo-

some. 120 connecting nodes were provided to the network produced by each chromosome gene list, resulting in minimum confidence scores of 0.933, 0.916 and 0.930 for 7A, 7B and 7D, respectively. The confidence score is a measure of the reliability of evidence for network interactions; therefore, the number of nodes applied and consistent confidence scores represent the optimal network generation for accurate comparison. Figure S1 displays the full gene networks representing each chromosome arm.

SNP density analysis and variation

Between 8.8× and 10.8× coverage of Illumina whole-genome paired read sequence data were generated for four Australian wheat cultivars Drysdale (173.11 Gbp), Gladius (185.60 Gbp), Excalibur (161.22 Gbp) and RAC875 (149.59 Gbp). The data were mapped to the three group 7 wheat chromosome assemblies as well as an assembly of chromosome arm 4AL (Hernandez *et al.*, 2012) to prevent the erroneous mapping of reads associated with 7BS/4AL translocation. Mapping was conducted with SOAP v2.21 (Li *et al.*, 2008) using the default parameters, allowing up to 2 mismatches per read. Only uniquely mapped paired reads were retained. The number of reads mapped from each dataset is presented in Table S5 and is consistent with the proportion of reads that could be expected to uniquely map with high stringency to the assembled portion of the genome.

SNPs were predicted using our in-house SGSautoSNP pipeline (Lorenec *et al.* 2012), which is available on request and based on the autoSNP algorithm (Barker *et al.*, 2003; Duran *et al.*, 2009). All predicted SNPs are presented on a GBrowse database at www.wheatgenome.info and summarized in Table 3 (full details in Table S6). SNP density was measured across each syntenic build using in-house scripts that count the number of SNPs occurring in a window of 50 kbp moving in a frame of 10 kbp (Figure 2). SNP density varied across each of the chromosomes, with a greater number of SNPs identified on chromosome 7A, and significantly fewer SNPs identified on chromosome 7D. Calculating the proportion of SNP transitions indicates the historic methylation profile of each subgenome (Table S6).

Acknowledgements

We thank Dr. Jarmila Číhalíková, Romana Šperková, Bc. and Ms. Zdeňka Dubská for their assistance with chromosome sorting. We also thank J. Perry Gustafson for valuable discussions on the evolutionary history of wheat. The authors would like to acknowledge funding support from Bioplatforms Australia, the Australian Research Council (Projects LP0882095, LP0883462 and DP0985953), the Department of Industry, Innovation, Science, Research and Tertiary Education (project CG120174), the Czech Science Foundation (grant No. P501/12/2554), and from the European Union (grant No. ED0007/01/01 Centre of the Region Haná for Biotechnological and Agricultural Research). Support also came from the Australian Genome Research Facility (AGRF), the Queensland Cyber Infrastructure Foundation (QCIF) and the Australian Partnership for Advanced Computing (APAC).

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**, 403–410.

- Barker, G., Batley, J., O' Sullivan, H., Edwards, K.J. and Edwards, D. (2003) Redundancy based detection of sequence polymorphisms in expressed sequence tag data using autoSNP. *Bioinformatics*, **19**, 421–422.
- Bekaert, M., Edger, P.P., Pires, J.C. and Conant, G.C. (2011) Two-phase resolution of polyploidy in the Arabidopsis metabolic network gives rise to relative and absolute dosage constraints. *Plant Cell*, **23**, 1719–1728.
- Berkman, P.J., Skarshewski, A., Lorenc, M., Lai, K., Duran, C., Ling, E.Y.S., Stiller, J., Smits, L., Imelfort, M., Manoli, S., McKenzie, M., Kubaláková, M., Šimková, H., Batley, J., Fleury, D., Doleel, J. and Edwards, D. (2011) Sequencing and assembly of low copy and genic regions of isolated *Triticum aestivum* chromosome arm 7DS. *Plant Biotechnol. J.* **9**, 768–775.
- Berkman, P.J., Skarshewski, A., Manoli, S., Lorenc, M.T., Stiller, J., Smits, L., Lai, K., Campbell, E., Kubaláková, M., Šimková, H., Batley, J., Doležel, J., Hernandez, P. and Edwards, D. (2012) Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation. *Theor. Appl. Genet.* **124**, 423–432.
- Caldwell, K.S., Dvorak, J., Lagudah, E.S., Akhunov, E., Luo, M.-C., Wolters, P. and Powell, W. (2004) Sequence Polymorphism in Polyploid Wheat and Their D-Genome Diploid Ancestor. *Genetics*, **167**, 941–947.
- Carollo, V., Matthews, D.E., Lazo, G.R., Blake, T.K., Hummel, D.D., Lui, N., Hane, D.L. and Anderson, O.D. (2005) GrainGenes 2.0: an improved resource for the small-grains community. *Plant Physiol.* **139**, 643–651.
- Chantret, N., Salse, J., Sabot, F., Rahman, S., Bellec, A., Laubin, B., Dubois, I., Dossat, C., Sourdil, P., Joudrier, P., Gautier, M.F., Cattolico, L., Beckert, M., Aubourg, S., Weissenbach, J., Caboche, M., Bernard, M., Leroy, P. and Chalhou, B. (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *Plant Cell*, **17**, 1033–1045.
- Chao, S., Zhang, W., Akhunov, E., Sherman, J., Ma, Y., Luo, M.-C. and Dubcovsky, J. (2009) Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L.) cultivars. *Mol. Breed.* **23**, 23–33.
- Dolores Vazquez, M., James Peterson, C., Riera-Lizarazu, O., Chen, X., Heesacker, A., Ammar, K., Crossa, J. and Mundt, C. (2012) Genetic analysis of adult plant, quantitative resistance to stripe rust in wheat cultivar 'Stephens' in multi-environment trials. *Theor. Appl. Genet.* **124**, 1–11.
- Dubcovsky, J. and Dvorak, J. (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science*, **316**, 1862–1866.
- Duran, C., Appleby, N., Clark, T., Wood, D., Imelfort, M., Batley, J. and Edwards, D. (2009) AutoSNPdb: an annotated single nucleotide polymorphism database for crop plants. *Nucleic Acids Res.* **37**, D951–953.
- Dvořák, J. (2009) Crops and Models. In: *Plant Genetics and Genomics* (Muehlbauer, G.M. and Feuillet, C., eds), pp. 685–711. New York: Springer.
- Dvorak, J., Luo, M.C., Yang, Z.L. and Zhang, H.B. (1998) The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. *Theor. Appl. Genet.* **97**, 657–670.
- Dvorak, J., Akhunov, E.D., Akhunov, A.R., Deal, K.R. and Luo, M.C. (2006) Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. *Mol. Biol. Evol.* **23**, 1386–1396.
- Eckardt, N.A. (2001) A Sense of Self: the Role of DNA Sequence Elimination in Allopolyploidization. *Plant Cell*, **13**, 1699–1704.
- Food and Agriculture Organisation of the United Nations (2012) *FAOSTAT Food Supply – Crops Primary Equivalent*.
- Freiling, M. and Thomas, B.C. (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res.* **16**, 805–814.
- Gegas, V.C., Nazari, A., Griffiths, S., Simmonds, J., Fish, L., Orford, S., Sayers, L., Doonan, J.H. and Snape, J.W. (2010) A genetic framework for grain size and shape variation in wheat. *Plant Cell*, **22**, 1046–1056.
- Genc, Y., Oldach, K., Verbyla, A., Lott, G., Hassan, M., Tester, M., Wallwork, H. and McDonald, G. (2010) Sodium exclusion QTL associated with improved seedling growth in bread wheat under salinity stress. *Theor. Appl. Genet.* **121**, 877–894.
- Giles, R.J. and Brown, T.A. (2006) GluDy allele variations in *Aegilops tauschii* and *Triticum aestivum*: implications for the origins of hexaploid wheats. *Theor. Appl. Genet.* **112**, 1563–1572.
- Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. and Rokhsar, D.S. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–1186.
- Hernandez, P., Martis, M., Dorado, G., Pfeifer, M., Galvez, S., Schaaf, S., Jouve, N., Šimková, H., Valárik, M., Doležel, J. and Mayer, K.F. (2012) Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *Plant J.* **69**, 377–386.
- Hossain, K.G., Kalavacharla, V., Lazo, G.R., Hegstad, J., Wentz, M.J., Kianian, P.M., Simons, K., Gehlhar, S., Rust, J.L., Syamala, R.R., Obeori, K., Bhamidimarri, S., Karunadharma, P., Chao, S., Anderson, O.D., Qi, L.L., Echalié, B., Gill, B.S., Linkiewicz, A.M., Ratnasiri, A., Dubcovsky, J., Akhunov, E.D., Dvorak, J., Miftahudin, Ross, K., Gustafson, J.P., Radhawa, H.S., Dilbirli, M., Gill, K.S., Peng, J.H., Lapitan, N.L., Greene, R.A., Bermudez-Kandianis, C.E., Sorrells, M.E., Feril, O., Pathan, M.S., Nguyen, H.T., Gonzalez-Hernandez, J.L., Conley, E.J., Anderson, J.A., Choi, D.W., Fenton, D., Close, T.J., McGuire, P.E., Qualset, C.O. and Kianian, S.F. (2004) A chromosome bin map of 2148 expressed sequence tag loci of wheat homoeologous group 7. *Genetics*, **168**, 687–699.
- Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R. and Gornicki, P. (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc. Natl. Acad. Sci. USA*, **99**, 8133–8138.
- Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57.
- Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13.
- International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.
- Kubaláková, M., Kovářová, P., Suchánková, P., Čihalíková, J., Bartoš, J., Lucretti, S., Watanabe, N., Kianian, S.F. and Doležel, J. (2005) Chromosome sorting in tetraploid wheat and its potential for genome analysis. *Genetics*, **170**, 823–829.
- Lai, K., Berkman, P.J., Lorenc, M.T., Duran, C., Smits, L., Manoli, S., Stiller, J. and Edwards, D. (2012) WheatGenome.info: an integrated database and portal for wheat genome information. *Plant Cell Physiol.* **52**, e2. (1–7).
- Li, R., Li, Y., Kristiansen, K. and Wang, J. (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics*, **24**, 713–714.
- Lorenc, M.T., Hayashi, S., Stiller, J., Lee, H., Manoli, S., Ruperao, P., Visendi, P., Berkman, P.J., Lai, K., Batley, J. and Edwards, D. (2012) Discovery of Single Nucleotide Polymorphisms in Complex Genomes Using SGAutoSNP. *Biology*, **1**, 370–382.
- Luo, M.C., Deal, K.R., Akhunov, E.D., Akhunova, A.R., Anderson, O.D., Anderson, J.A., Blake, N., Clegg, M.T., Coleman-Derr, D., Conley, E.J., Crossman, C.C., Dubcovsky, J., Gill, B.S., Gu, Y.Q., Hadam, J., Heo, H.Y., Huo, N., Lazo, G., Ma, Y., Matthews, D.E., McGuire, P.E., Morrill, P.L., Qualset, C.O., Renfro, J., Tabanao, D., Talbert, L.E., Tian, C., Toleno, D.M., Warburton, M.L., You, F.M., Zhang, W. and Dvorak, J. (2009) Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in *Triticeae*. *Proc. Natl. Acad. Sci. USA* **106**, 15780–15785.
- Matthews, D.E., Carollo, V.L., Lazo, G.R. and Anderson, O.D. (2003) GrainGenes, the genome database for small-grain crops. *Nucleic Acids Res.* **31**, 183–186.
- Mayer, K.F., Taudien, S., Martis, M., Šimková, H., Suchánková, P., Gundlach, H., Wicker, T., Petzold, A., Felder, M., Steuernagel, B., Scholz, U., Graner, A., Platzer, M., Doležel, J. and Stein, N. (2009) Gene content and virtual gene order of barley chromosome 1H. *Plant Physiol.* **151**, 496–505.
- McFadden, E.S. and Sears, E.R. (1946) The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *J. Hered.* **37**, 81–107.
- Nesbitt, M. and Samuel, D. (1995) From staple crop to extinction? The archaeology and history of the hulled wheats. In: *First International Workshop on Hulled Wheats* (Padulosi, S., Hammer, K. and Heller, J., eds), pp. 40–99. Castelvecchio Pascoli, Tuscany, Italy: Biodiversity International.

- Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Felts, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Ollilar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob ur, R., Ware, D., Westhoff, P., Mayer, K.F., Messing, J. and Rokhsar, D.S. (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, **457**, 551–556.
- Šafář, J., Šimková, H., Kubaláková, M., Čihalíková, J., Suchánková, P., Bartoš, J. and Doležel, J. (2010) Development of chromosome-specific BAC resources for genomics of bread wheat. *Cytogenet. Genome Res.* **129**, 211–223.
- Salamini, F., Ozkan, H., Brandolini, A., Schafer-Pregl, R. and Martin, W. (2002) Genetics and geography of wild cereal domestication in the near east. *Nat. Rev. Genet.* **3**, 429–441.
- Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Kane, T., Thane, N.C., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golsner, W., Kim, H., Lee, S., Lin, J. and Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambrose, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.P., Barbazuk, W.B., Baucum, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.M., Deragon, J.M., Estill, J.C., Fu, Y., Jeddeloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A. and Wilson, R.K. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115.
- Schnurbusch, T., Collins, N.C., Eastwood, R.F., Sutton, T., Jefferies, S.P. and Langridge, P. (2007) Fine mapping and targeted SNP survey using rice-wheat gene colinearity in the region of the Bo1 boron toxicity tolerance locus of bread wheat. *Theor. Appl. Genet.* **115**, 451–461.
- Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguéz, P., Doerks, T., Stark, M., Müller, J., Bork, P., Jensen, L.J. and von Mering, C. (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, D561–D568.
- Talbert, L.E., Smith, L.Y. and Blake, N.K. (1998) More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. *Genome*, **41**, 402–407.
- Thomas, B.C., Pedersen, B. and Freeling, M. (2006) Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res.* **16**, 934–946.
- Vogel, J.P., Garvin, D.F., Mockler, T.C., Schmutz, J., Rokhsar, D., Bevan, M.W., Barry, K., Lucas, S., Harmon-Smith, M., Lail, K., Tice, H., Schmutz Leader, J., Grimwood, J., McKenzie, N., Huo, N., Gu, Y.Q., Lazo, G.R., Anderson, O.D., Vogel Leader, J.P., You, F.M., Luo, M.C., Dvorak, J., Wright, J., Febrer, M., Idziak, D., Hasterok, R., Lindquist, E., Wang, M., Fox, S.E., Priest, H.D., Filichkin, S.A., Givan, S.A., Bryant, D.W., Chang, J.H., Mockler Leader, T.C., Wu, H., Wu, W., Hsia, A.P., Schnable, P.S., Kalyanaraman, A., Barbazuk, B., Michael, T.P., Hazen, S.P., Bragg, J.N., Laudencia-Chingcuanco, D., Weng, Y., Haberer, G., Spannagl, M., Mayer Leader, K., Rattei, T., Mitros, T., Lee, S.J., Rose, J.K., Mueller, L.A., York, T.L., Wicker Leader, T., Buchmann, J.P., Tanskanen, J., Schulman Leader, A.H., Gundlach, H., Bevan, M., Costa de Oliveira, A., da, C.M.L., Belknap, W., Jiang, N., Lai, J., Zhu, L., Ma, J., Sun, C., Pritham, E., Salse Leader, J., Murat, F., Abrouk, M., Mayer, K., Bruggmann, R., Messing, J., Fahlgren, N., Sullivan, C.M., Carrington, J.C., Chapman, E.J., May, G.D. and Zhai, J., Ganssmann, M., Guna Ranjan Gurazada, S., German, M., Meyers, B.C., Green Leader, P.J., Tyler, L., Wu, J., Thomson, J., Chen, S., Scheller, H.V., Harholt, J., Ulvskov, P., Kimbrel, J.A., Bartley, L.E., Cao, P., Jung, K.H., Sharma, M.K., Vega-Sanchez, M., Ronald, P., Dardick, C.D., De Bodt, S., Verelst, W., Inze, D., Heese, M., Schnittger, A., Yang, X., Kalluri, U.C., Tuskan, G.A., Hua, Z., Vierstra, R.D., Cui, Y., Ouyang, S., Sun, Q., Liu, Z., Yilmaz, A., Grotewold, E., Sibout, R., Hematy, K., Mouille, G., Hofte, H., Michael, T., Pelloux, J., O'Connor, D., Schnable, J., Rowe, S., Harmon, F., Cass, C.L., Sedbrook, J.C., Byrne, M.E., Walsh, S., Higgins, J., Li, P., Brutnell, T., Unver, T., Budak, H., Belcram, H., Charles, M., Chalhouh, B. and Baxter, I. (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, **463**, 763–768.
- Vrána, J., Kubaláková, M., Šimková, H., Čihalíková, J., Lysák, M.A. and Doležel, J. (2000) Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). *Genetics*, **156**, 2033–2041.
- Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I., Cheng, F., Huang, S., Li, X., Hua, W., Freeling, M., Pires, J.C., Paterson, A.H., Chalhouh, B., Wang, B., Hayward, A., Sharpe, A.G., Park, B.S., Weissshaar, B., Liu, B., Li, B., Tong, C., Song, C., Duran, C., Peng, C., Geng, C., Koh, C., Lin, C., Edwards, D., Mu, D., Shen, D., Soumpourou, E., Li, F., Fraser, F., Conant, G., Lassalle, G., King, G.J., Bonnema, G., Tang, H., Belcram, H., Zhou, H., Hirakawa, H., Abe, H., Guo, H., Jin, H., Parkin, I.A., Batley, J., Kim, J.S., Just, J., Li, J., Xu, J., Deng, J., Kim, J.A., Yu, J., Meng, J., Min, J., Poulain, J., Hatakeyama, K., Wu, K., Wang, L., Fang, L., Trick, M., Links, M.G., Zhao, M., Jin, M., Ramchiary, N., Drou, N., Berkman, P.J., Cai, Q., Huang, Q., Li, R., Tabata, S., Cheng, S., Zhang, S., Sato, S., Sun, S., Kwon, S.J., Choi, S.R., Lee, T.H., Fan, W., Zhao, X., Tan, X., Xu, X., Wang, Y., Qiu, Y., Yin, Y., Li, Y., Du, Y., Liao, Y., Lim, Y., Narusaka, Y., Wang, Z., Li, Z., Xiong, Z. and Zhang, Z. (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* **43**, 1035–1039.
- Wanjugi, H., Coleman-Derr, D., Huo, N., Kianian, S.F., Luo, M.-C., Wu, J., Anderson, O. and Gu, Y.Q. (2009) Rapid development of PCR-based genome-specific repetitive DNA junction markers in wheat. *Genome*, **52**, 576–587.
- Wicker, T., Buchmann, J.P. and Keller, B. (2010) Patching gaps in plant genomes results in gene movement and erosion of colinearity. *Genome Res.* **20**, 1229–1237.
- Wicker, T., Mayer, K.F.X., Gundlach, H., Martis, M., Steuernagel, B., Scholz, U., Šimková, H., Kubaláková, M., Choulet, F., Taudien, S., Platzer, M., Feuillet, C., Fahima, T., Budak, H., Doležel, J., Keller, B. and Stein, N. (2011) Frequent Gene Movement and Pseudogene Evolution Is Common to the Large and Complex Genomes of Wheat, Barley, and Their Relatives. *Plant Cell*, **23**, 1706–1718.
- Woodhouse, M.R., Schnable, J.C., Pedersen, B.S., Lyons, E., Lisch, D., Subramaniam, S. and Freeling, M. (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. *PLoS Biol.* **8**, e1000409.
- Zerbino, D.R. and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829.

Supporting information

Additional Supporting information may be found in the online version of this article:

Figure S1 Network image produced by STRING representing the scale of gene networking on 7A (A), 7B (B), and 7D (C).

Figure S2 Histogram displaying the number of bin-mapped EST sequences across all wheat chromosomes that have aligned to the assemblies of 7A (red), 7B (blue), and 7D (green).

Figure S3 CMap image representing the comparison of genetic markers from *A. tauschii* against the 7D syntenic build.

Table S1 Full summary of Syntenic Build statistics.

Table S2 GenomeZipper of wheat group 7 chromosomes against

B. distachyon, *O. sativa*, and *S. bicolor* in both syntenic (A) and nonsyntenic (B) regions.

Table S3 STRING GO terms enrichment analysis with all group 7 genes used as the background.

Table S4 Table listing chromosome-arm specific bin-mapped EST RBB alignment results for contigs assembled from each chromosome arm.

Table S5 Read mapping statistics for Illumina 100 bp paired-end reads from 4 Australian varieties (Gladius, Drysdale, Excalibur, RAC875).

Table S6 SNPs called between 4 Australian wheat cultivars (Gladius, Drysdale, Excalibur, RAC875).