Supporting Information

Luo et al. 10.1073/pnas.1219082110

SI Text

Genetic Map Construction. *Aegilops tauschii* ssp. *strangulata* accession AL8/78 was collected by V. Jaaska (Department of Botany, Institute of Zoology and Botany, Tartu, Estonia) in Yerevan, Armenia, near the Hrazdan River. Accession AS75 (*Ae. tauschii* ssp. *typica*) was collected in Xi'an, Shaanxi Province, China. The former accession was used for the construction of large-insert libraries, and both were used as the parents of a biparental mapping population for the construction of a genetic map.

Earlier, we discovered 195,631 genic SNPs between accessions AL8/78 and AS75 with the genomewide SNP discovery pipeline AGSNP (1). About 84% of the SNPs were real; the rest were sequencing errors (1). To construct a 10K Infinium iSelect SNP array, we selected the best SNPs present in sequence contigs reported by You at al (1). The Infinium II type SNPs were then selected from this pool to maximize the number of SNP assays in the 10K Infinium and SNP genotyping performance.

We submitted the SNPs to Illumina for evaluation using *Illumina's Assay Design Tool* (*ADT*). On the basis of the ADT design scores, we submitted 10,000 high-score SNPs for manufacturing and ended up with 9,485 functional assays in the 10K Infinium SNP array. This population of SNP assays included 515 SNPs located in wheat expressed sequence tags (ESTs) (labeled by GenBank accession numbers) (2) that have previously been used for an Illumina GoldenGate SNP assay array construction (3). We used these SNPs to align the Infinium *Ae. tauschii* map with the preceding *Ae. tauschii* GoldenGate map (3).

We grew 1,102 AL8/78 × AS75 F_2 plants in the greenhouse and extracted DNA from isolated nuclei as described earlier (4). The UC Davis Genome Center used 3 μg of genomic DNA (300 ng/μL) per plant for performing Infinium SNP genotyping assays using protocols provided by the Infinium manufacturer (Illumina). We processed the 10K Infinium genotyping data with the GenomeStudio software (Illumina) and manually examined the graphs of genotyped DNAs of the F_2 plants and the AL8/78 and AS75 parental controls for clustering. If an SNP assay performed well, we expected three well-separated clusters of genotype scores in the 1:2:1 codominant monohybrid ratio, as illustrated in Fig. S1. The clustering of 1,214 SNP markers (14.5%) was inadequate, and they were excluded. The remaining 7,185 SNP assays generated genotype clustering similar to that shown in the upper panel of Fig. S1 and were used for the construction of the genetic map.

We generated a genotype matrix for the 1,102 *Ae. tauschii* F_2 plants from the GenomeStudio SNP genotyping score output and used it as an input in the MultiPoint mapping software (5) using the following settings: cluster threshold (recombination rate) of 0.1, Jackknife value 90, number of iteration 10, and Kosambi function. We obtained seven linkage groups, one for each of the seven chromosomes of the *Ae. tauschii* genome. We manually examined the marker order and rechecked the matrix data and GenomeStudio clustering for markers showing low confidence order on the maps. We executed three iterations of map construction and checked the matrix and GenomeStudio data each time. Some groups of markers showed no recombination within the groups, and we based the order of those markers on synteny comparisons with *Brachypodium distachyon*, rice (*Oryza sativa*), and sorghum (*Sorghum bicolor*) and their location in bacterial artificial chromosome (BAC) contigs.

BAC Libraries, BAC Clone Fingerprinting, Fingerprint Editing, and BAC Contig Assembly. We confirmed the identity of greenhouse-grown AL8/78 plants targeted for the construction of BAC libraries by sequencing the PCR amplicons of ESTs AJ603554, BE403345, BE488719, BE518031, and BG263347 that had been sequenced previously in a number of *Ae. tauschii* accessions including AL8/78 (3, 6).We collected leaves from the plants, immediately froze them in liquid nitrogen, placed them into plastic bags, and mailed them to Amplicon Express on dry ice for BAC library construction. Amplicon Express constructed the BamHI, EcoRI, HindIII, and MboI BAC libraries using the pCC1 BAC vector (EPICENTRE) and DH10B host cells. The total number of clones per library, average insert sizes, number of clones used for fingerprinting, and other characteristics of the libraries are summarized in Table S1 (first four rows). Inserts in 100 clones per library were sized with pulse-field electrophoresis.

A total of 406,944 BAC clones from these four libraries (Table S1) were fingerprinted with a SNaPshot high-information content fingerprinting (HICF) method described by Luo et al. (7) and modified by Gu et al. (8). We also randomly selected 22,810 clones from a HindIII BAC library of AL8/78 previously constructed and fingerprinted with a technique described by Luo et al. (7). Clones from that earlier library construction and assembly (9) will be called phase I clones to distinguish them from the BAC clones produced here (phase II clones; Table S1). We refingerprinted these 22,810 phase I clones with the modified fingerprinting method used here and included them into the present contig assembly for future alignment of phase I contigs with the phase II contigs. We also refingerprinted 31,805 BAC and binary BAC (BiBAC, vector pCLD04541) clones located near the ends of phase I contigs for the same purpose as the HindIII BAC clones (Table S1). Unfortunately, the identity of these clones was incorrect, and they could not be used for associating phase I contigs with phase II contigs. However, because they were already fingerprinted, we included them into the phase II contig assembly.

In total, we fingerprinted 461,706 clones of an average insert size length of 120.5 kb (Table S1). The fingerprints were edited with the FPMiner software (8) using the default settings. During fingerprint editing, we retained restriction fragments only in the 70- to 1,000-bp size range, excluded vector fragments, clones failing fingerprinting or lacking inserts, and clones with less than 30 or more than 220 fragments, and removed cross-contaminated samples using a module in the FPMiner. Cross-contamination was detected as clones residing in neighboring wells and sharing 30% or more of the mean number of fragments in their profiles. After editing, we were left with 399,448 fingerprints for contig assembly (Table S1). The average insert length of the edited clones is unknown but must be >120.5 kb because clones with short inserts were removed from the pool of fingerprinted clones during the fingerprint editing phase.

We assembled the fingerprinted clones into contigs with FPC software (version 9.3, www.agcol.arizona.edu/software/fpc/) using the following strategy. We set the tolerance at 5 (=0.5 bp) throughout the assembly. We performed the initial assembly at Sulston cutoff of 1×10^{-70}, which was followed by several rounds of DQering, until all contigs contined <15% questionable (Q) clones. We then reduced Sulston cutoff stringency and performed end-to-end and singleton-to-end contig merges, requiring two or more clones per merge. Sulston cutoff stringency reduction and contig merging were repeated until a Sulston score of 1×10^{-22} was reached, at which point the assembly was terminated. The assembly resulted in a total of 3,153 contigs and 15,683 singletons.

BAC Pool Construction, Genotyping, and Deconvolution. We used a 5-D pooling strategy for BAC contig anchoring (10, 11). We isolated BAC DNA with the R.E.A.L. kit (Qiagen) and used 5 μL each (20 μL total) for a column and row pools in a stack of 100 plates. We produced eight separate sets of row/column pools, 1× genome coverage each, to minimize the number of false-positive pool intersections. To generate plate pools, we inoculated 100 mL of LB medium in a 250-mL flask with all 384 clones in a single plate, grew them overnight, and isolated DNA with a standard alkaline-lysis protocol. We arranged DNAs of plate-pools into an 18 × 18 grid, and combined aliquots into column and row superpools. We generated four sets of superpools, each was ~3× genome equivalent. We added 300 ng of AS75 genomic DNA to all BAC pools to preempt nonspecific PCR amplification in the absence of a target DNA. A total of 10 μg of DNA per pool was submitted to the UC Davis Genome Center for genotyping with the 10K *Ae. tauschii* Infinium array.

The term deconvolution means the identification of the positive BAC clone(s) among the clones forming the 5-D BAC pools. Because the deconvolution program and its algorithms have been published (11), we will describe here only the basic idea and focus on details by which we implemented BAC pool deconvolution. The deconvolution program identified intersections between positive BAC row-pools, column-pools, and plate-pools from the 10K Infinium assay resulting in some false-positive intersections. To eliminate the need for PCR discrimination between true-positive and false-positive intersections, we used the distribution of BAC clones in contigs to discriminate between true-positive and false-positive intersections (11). BAC clones that are true positives must be neighbors in a contig and overlap, whereas clones that are false positive are distributed randomly among contigs. Only contigs that were anchored at a locus on the genetic map by two or more overlapping BAC clones were therefore accepted as true positive clones.

Because of the variable amounts of BAC DNAs in the pools, 10K Infinium BAC pool genotyping data failed to produce the clear-cut clustering seen in the upper panel in Fig. S1. Instead, we obtained diffuse plots only vaguely resembling the expected clusters (lower panel in Fig. S1). The fact that at the same time the genomic DNAs of AL8/78 and AS75 produced tight clusters near the *x* and *y* axes convinced us that the Infinium assay performed well. To determine where in the plots the negative and positive BAC pools clustered, we manually deconvoluted the 10K Infinium BAC pool genotyping data for 34 markers on the 1D genetic map. Negative BAC pools (sharing the genotype with AS75) were located in a single tight cluster near one of the ordinates together with the AS75 genomic DNAs (green dots in the green oval in the lower panel of Fig. S1. The positive BAC pools were located in the diffused cloud of dots along the ordinate containing the genomic DNAs showing the AL8/78 nucleotide (red dots in the red oval). We empirically determined that all dots with fluorescence 1.5 times the background fluorescence of AS75 genomic DNAs (fluorescence A in the lower panel of Fig. S1) is a robust boundary separating the positive BAC pools from negative BAC pools. Thus, all pools within the blue rectangle in the lower panel of Fig. S1 are positive BAC pools and superpools.

Manual Contig Editing and Physical Map Construction. The purpose of manual contig editing was to detect chimeric contigs and dissociate them. We examined the genetic map location of markers integrated into each contig. If the markers were in two separate regions on the genetic map, the contig was deemed chimeric and was manually disjoined using FPC tools. As illustrated in Fig. S2, a false join caused by a chimeric BAC clone can be easily detected on the FPC's CB map of a contig. In addition to the diagnostic pattern, the clone causing a false join is almost always a Q clone (Fig. S2). We examined CB maps of all anchored contigs for false joins and of unanchored contigs >1 Mb, and clones causing false joins were removed, which separated each chimeric contig into two. We also coassembled *Ae. tauschii* contigs with BAC clones from subgenomic BAC libraries (Fig. S3). Luo et al. (12) showed that contig coassembly using subgenomic BAC libraries fingerprinted with the SNaPshot HICF technique is an effective strategy for detecting BAC clone relationships. Here we used this technique for detecting *Ae. tauschii* chimeric contigs consisting of clones from different *Ae. tauschii* chromosomes or chromosome arms. We coassembled *Ae. tauschii* contigs with fingerprinted clones from subgenomic BAC libraries constructed from DNA isolated from the following flow-sorted wheat cv Chinese Spring chromosomes or chromosome arms: 30,067 fingerprinted and edited clones from a 1D-4D-6D BAC library (13), 30,157 fingerprinted and edited clones from a 3DS BAC library (12), 39,852 fingerprinted and edited clones from a 7DS BAC library (14), and 43,492 fingerprinted clones from a 7DL BAC library (14). BAC libraries of wheat chromosomes 2D and 5D and chromosome arm 3DL fingerprinted with a technique identical to that used here were not available to us.

Combined information provided by the contig marker anchoring, contig CB map, and contig coassembly was sufficiently redundant to detect and disjoin most of the chimeric contigs. Disjoining of chimeric contigs increased the total number of contigs from 3,153 to 3,578 and decreased their average length from 1,509 to 1,339 kb.

Extension of Marker Sequences. We extended the sequences containing SNP markers on the genetic and physical maps with 3.1× genome equivalent of Roche 454 WGS sequences and assembled sequence contigs. The Roche 454 sequence contigs were then stepwise extended with 50× genome equivalent of short Illumina contigs.

Roche 454 genomic library construction and sequencing. We prepared and sequenced the 454 sequencing library according to the manufacturer's instructions (GS FLX Titanium General Library preparation kit/emPCRkit sequencing kit; Roche Diagnostics). Briefly, we sheared 10 μg of *Ae. tauschii* accession AL78/78 genomic DNA by nebulization and fractionated it with agarose gel electrophoresis to isolate 400- to 750-bp fragments and used the sized fragments to construct a single-stranded shotgun library. We quantified the library by fluorometry using the Quant-iT RiboGreen reagent and processed it by emulsion PCR amplification. We sequenced the library with GS FLX Titanium following the manufacturer's recommendations (Roche Diagnostics).

Illumina library barcoding and sequencing. We quantified *Ae. tauschii* genomic DNA using the Qubit flourometer and used ~2 μg of DNA for the construction of the standard 300-bp and overlapping 180-bp Illumina libraries. We sheared the DNA by adaptive focused acoustics (using the Covaris instrument) and end-repaired it using T4 DNA polymerase, Klenow fragment, and T4 polynucleotide kinase. To add a single 3′ deoxyA overhang, we treated fragments with Klenow fragment (3′–5′ exonuclease) and ligated them to standard paired-end Illumina adapters. Qiagen columns were used for purification between steps. For the 300-bp library, we size-selected the fragments in the range of 350–450 bp using agarose gel electrophoresis, and for the 180-bp overlapping library, we size-selected the fragments for an insert size in the range of 190–210 bp using the Caliper Labchip XT instrument. We then PCR amplified each library using Phusion DNA polymerase in HF buffer for 12 cycles and quantified using the Agilent BioAnalyzer.

To construct mate-pair libraries (2 and 5 Kb), we sheared 10 μg of genomic DNA with Covaris, end-repaired it using T4 DNA polymerase, Klenow fragment, and T4 polynucleotide kinase, and added Biotinylated bases using T4 DNA polymerase, Klenow fragment, and T4 polynucleotide kinase. Qiagen beads were used for purification between steps. We size-selected DNA

fragments in the 2- or 5-Kb range using agarose gel electrophoresis and quantitated them with the Agilent BioAnalyzer. We circularized DNA fragments with ligase, digested linear DNA with exonuclease, fragmented the circular DNA with the Covaris to ~400 bp, and selected biotinylated fragments by binding to streptavidin magnetic beads. We repaired biotinylated fragments as above, A-tailed them using Klenow, and ligated them to standard Illumina adapters. Each library was then PCR amplified using Phusion DNApolymerase in HF buffer for 18 cycles. We performed final size selection for 350- to 650-bp fragments by agarose gel electrophoresis and quantified using the Agilent BioAnalyzer. All libraries were normalized to 10 nM before loading on the Illumina sequencers.

Illumina sequencing. We sequenced the 300- and 180-bp standard *Ae. tauschii* libraries with 100-bp paired end read lengths and the 2- and 5-Kb mate pairs using paired-end 50-bp read lengths, using Illumina GAIIx or HiSeq2000 instruments with paired-end modules. A 1% phiX control library was spiked into each sample lane to aid in quality monitoring as the runs progressed. We used the most current versions of the Illumina instrument control software and the Illumina flowcell and reagent kits available at the time each library was sequenced.

Initial Illumina sequence analysis. We processed images generated on Illumina GAIIx or HiSeq2000 sequencers and performed base-calling on the fly using the Illumina Real Time Analysis (RTA) software. The files were then transferred to a secondary Linux server for further processing. The .bcl files produced by the RTA software on the instrument contained base-call and quality score information in binary format. The .bcl files were converted to FASTQ format by the CASAVA pipeline (v1.7/v1.8), which also provided run summary and quality information. Illumina FASTQ files were then uploaded to the NCBI short read archive (SRA). See www.cshl.edu/genome/wheat for SRA accession information.

Roche and Illumina contig assembly. We used 3.1× *Ae. tauschii* genome equivalents of Roche 454 reads for de novo assembly of contigs with the Roche gsAssembler using default settings. The assembly generated 1,070,122 contigs of a total length of 584,671,146 bp and an N50 of 835 bp. To filter out repetitive DNA, we searched homology between the 454 contigs and the TREP database (a curated database of repeat elements in the tribe Triticeae; http://wheat.pw.usda.gov/ITMI/Repeats/).

We performed a similar manipulation with the 50× Illumina reads. They too were filtered by homology search against the TREP database. By filtering out repeated sequences, we reduced the Illumina sequences from the original 2,566,522,820 to 1,518,407,964 bp (a 59.2% reduction). We assembled the filtered Illumina reads with Velvet, but because of limited computer memory, we were able to use fragment reads and 300-bp paired-end data only up to a 3.5× *Ae. tauschii* genome equivalent. We performed 17 such independent assemblies, which on average assembled 4.8 million sequences with an average N50 value of 221 bp. A total of 714 Mb of genomic DNA was assembled into these short contigs.

Contig extension. The 454 contigs constructed from the 3.1× Roche 454 reads were extended with 50× Illumina contigs. First we performed a blastN search against the Velvet Illumina read assembly dataset to identify the Illumina contigs corresponding to the 454 reads. We then stepwise extended the 454 contigs by using 100 bp from the end of each contig in a blast search against the Illumina sequences to extract reads or contigs at an *E*-value of −30. We repeated this step until an attempt to extend a contig failed due to the absence of reads matching the end sequence or due to the end sequence matching a repetitive sequence. The average length of the 7,185 contigs containing SNP markers was extended to 7,869 bp, with a total cumulative length of 61 Mb and an N50 of 10,830 bp. The extended marker sequence length ranged from 348 to 54,605 bp.

We used a genome annotation pipeline MAKER (http://gmod.org/wiki/MAKER) for sequence annotation to generate a set of ab initio gene predictions in the 7,185 extended contigs. MAKER identified repeats using the TREP database, aligned ESTs and protein sequences with contig sequences, produced ab initio gene predictions, and automatically synthesized these data into gene annotation classes with evidence-based quality indices. We aligned wheat and barley full-length ESTs and assembled wheat EST contigs (http://plantta.jcvi.org/cgi-bin/plantta_release.pl) with the extended sequence contigs in the MAKER pipeline. We also used plant protein datasets for homology comparison in gene prediction. In total we predicted 17,093 protein-encoding genes or gene fragments in the 7,185 extended sequence contigs. We assumed that 9,716 of these genes that were without any gap in the coding sequence and the predicted gene sequences were fully aligned with ESTs or annotated proteins were complete genes. MAKER also provided a list of sequences that showed partial alignment to ESTs or proteins. In this case, sequences partially matching ESTs were named as gene fragments_EST and those that did not match any wheat or barley EST but partially matched sequences in the nonredundant protein database were named gene fragments_protein. They could be pseudogenes or novel genes not present in databases. We also calculated the average gene, exon, and intron lengths for the 9,716 annotated genes.

The output of MAKER was used to create a gff file and used in our Gbrowse web interface build (http://probes.pw.usda.gov/cgi-bin/gb2/gbrowse/wheat_D_marker/). A spreadsheet of the 17,093 genes and gene fragments including name, location on the genetic map, locations of homologous genes in *B. distachyon*, rice, and sorghum and gene ontology (GO) is at http://probes.pw.usda.gov/WheatDMarker/downloads/GeneList.xls.

Dot Plots. To make dot plots as shown in Fig. 2 *B* and *C*, and Fig. S4, we aligned *Ae. tauschii* marker fasta sequences to annotated proteins of reference grass species using NCBI BLASTX. To increase sensitivity, we predicted translated sequences of *Ae. tauschii* markers on the basis of FGENESH (PMID:10779491) and aligned them to annotated reference proteins by BLASTP. We determined the collinear relationships between best significantly aligned marker and reference genes (*E*-value $\leq 1E^{-10}$) using DAGchainer (PMID:15247098), and if necessary, filtered paralogous chromosomal relationships to leave just orthologous collinear gene pairs. We collapsed duplicate loci so that each marker or gene was represented exactly once. We graphed marker and reference gene loci on the basis of their rank position along chromosomes. For plots between rice, sorghum, and *B. distachyon*, collinear genes were detected among orthologous genes as classified in Gramene Release 32 (November 2010) on the basis of Compara phylogenetic trees (PMID: 19029536; PMID: 21076153). The following reference genome annotations were used: *Oryza sativa*, MSU6.1; *Brachypodium distachyon*, JGI Brachy1.2; and *Sorghum bicolor,* JGI Sbi1.4 (PMID: 17145706, PMID: 20148030, and PMID: 19189423, respectively).

To make dot plots shown in Fig. S6, we first performed BLASTX analysis of the 17,093 genes and gene fragments against the annotated rice genome. The top rice hit of each *Ae. tauschii* gene or gene fragments was recorded with its coordinate in the rice genome. The dot plots of individual *Ae. tauschii* chromosomes were graphed by plotting each marker locus along the physical map against the corresponding top hit in the rice genome.

1. You FM, et al. (2011) Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence. *BMC Genomics* 12:59.

2. Qi LL, et al. (2004) A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. *Genetics* 168(2): 701–712.

3. Luo MC, et al. (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(37):15780–15785.
4. Dvorak J, McGuire PE, Cassidy B (1988) Apparent sources of the A genomes of wheats inferred from the polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. *Genome* 30(5):680–689.
5. Mester DI, et al. (2006) Multilocus consensus genetic maps (MCGM): formulation, algorithms, and results. *Comput Biol Chem* 30(1):12–20.
6. Akhunov ED, et al. (2010) Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics* 11:702.
7. Luo MC, et al. (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the SNaPshot labeling kit and sizing of restriction fragments by capillary electrophoresis. *Genomics* 82(3):378–389.
8. Gu YQ, et al. (2009) A BAC-based physical map of *Brachypodium distachyon* and its comparative analysis with rice and wheat. *BMC Genomics* 10:496.
9. Luo MC, et al. (2003) Construction of contigs of *Ae. tauschii* genomic DNA fragments cloned in BAC and BiBAC vectors. *Proceedings of the Tenth International Wheat Genetics Symposium*, eds Pogna NE, Romano M, Pogna EA, Galterio G (S.IM.I., Rome, Italy), pp 293–296.
10. Luo MC, et al. (2009) A high-throughput strategy for screening of bacterial artificial chromosome libraries and anchoring of clones on a genetic map constructed with single nucleotide polymorphisms. *BMC Genomics* 10:28.
11. You FM, et al. (2010) A new implementation of high-throughput five-dimensional clone pooling strategy for BAC library screening. *BMC Genomics* 11:692.
12. Luo MC, et al. (2010) Feasibility of physical map construction from fingerprinted bacterial artificial chromosome libraries of polyploid plant species. *BMC Genomics* 11:122.
13. Janda J, et al. (2004) Construction of a subgenomic BAC library specific for chromosomes 1D, 4D and 6D of hexaploid wheat. *Theor Appl Genet* 109(7): 1337–1345.
14. Šimková H, et al. (2011) BAC libraries from wheat chromosome 7D: Efficient tool for positional cloning of aphid resistance genes. *J Biomed Biotechnol* 2011:302543.

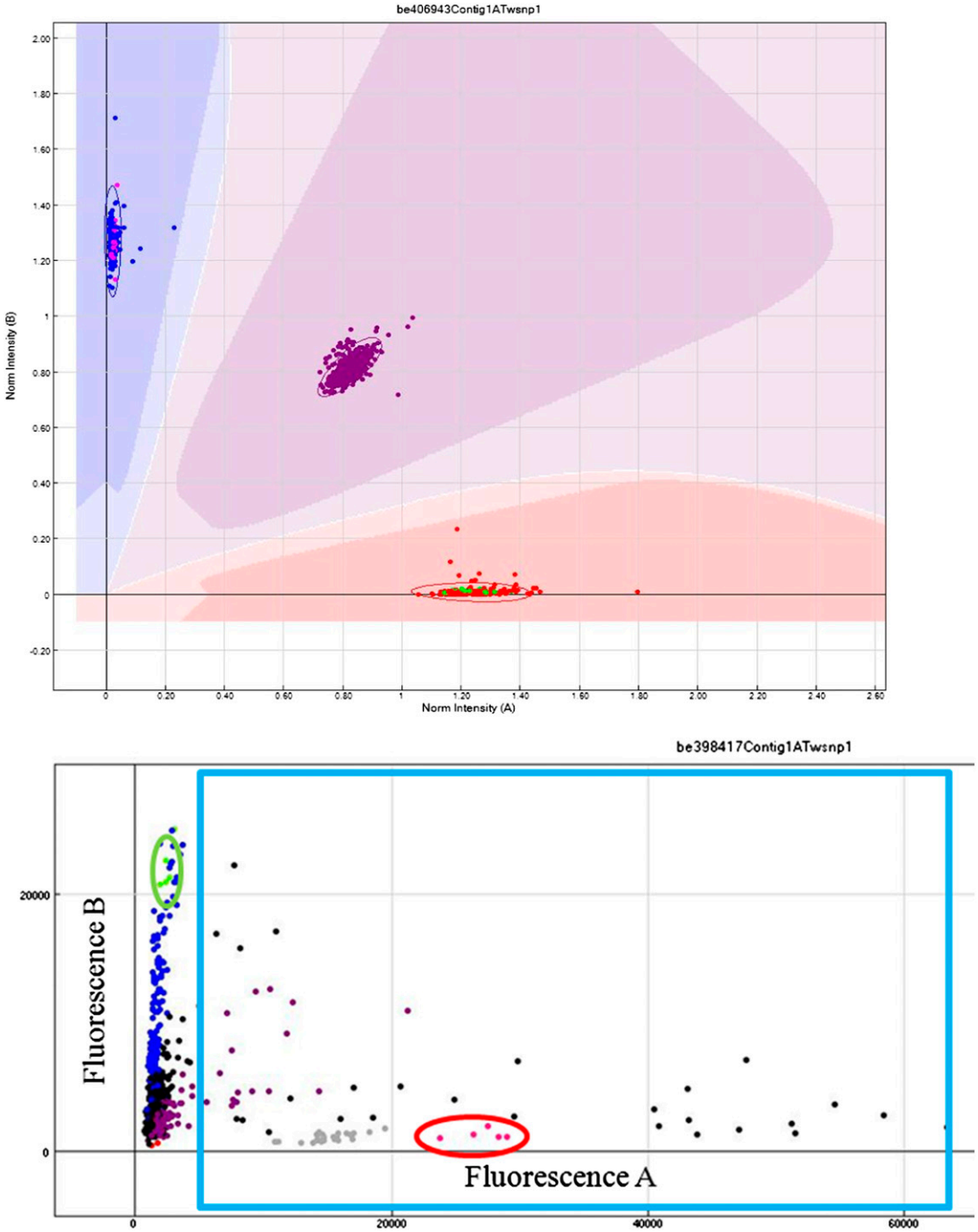

Fig. S1. Comparison of genotype clustering in Cartesian graphs in GenomeStudio outputs for F_2 plants of the mapping population (*Upper*) and 5-D BAC pools (*Lower*). *Upper* shows 1,102 *Ae. tauschii* F_2 plants from the cross AS75 × AL8/78 genotyped with an Infinium SNP assay for an SNP in EST locus BE406943. The red and blue clusters contain homozygous genotypes and the purple cluster contains heterozygous genotypes. *Lower* is a GenomeStudio Cartesian graph from the genotyping of the five-dimensional BAC pools of *Ae. tauschii* accession AL8/78 with the Infinium SNP assay BE398417Contig1ATwsnp1. We added a constant amount of *Ae. tauschii* AS75 genomic DNA to each BAC pool to preempt nonspecific PCR amplification in the assay. Each dot in the graph represents a DNA sample, either a BAC pool or genomic DNA of accession AS75 (green dots in the green oval) and accession AL8/78 (red dots in the red oval). The horizontal and vertical coordinates of a dot are quantified fluorescence A and B, which in this specific case measures the amount of AL8/78 nucleotide and AS75 nucleotide in a DNA sample, respectively. Note that only the AL8/78 and AS75 generated clusters reminiscent the clusters in the upper panel. We empirically determined that pools with <1.5 times the average fluorescence of the AS75 genomic DNAs in the AL8/78 nucleotide fluorescence spectrum (fluorescence A in this case) were likely negative. Those showing >1.5 times the average fluorescence of the AS75 genomic DNA (BAC pools and genomic DNAs in the blue box) were likely positive.

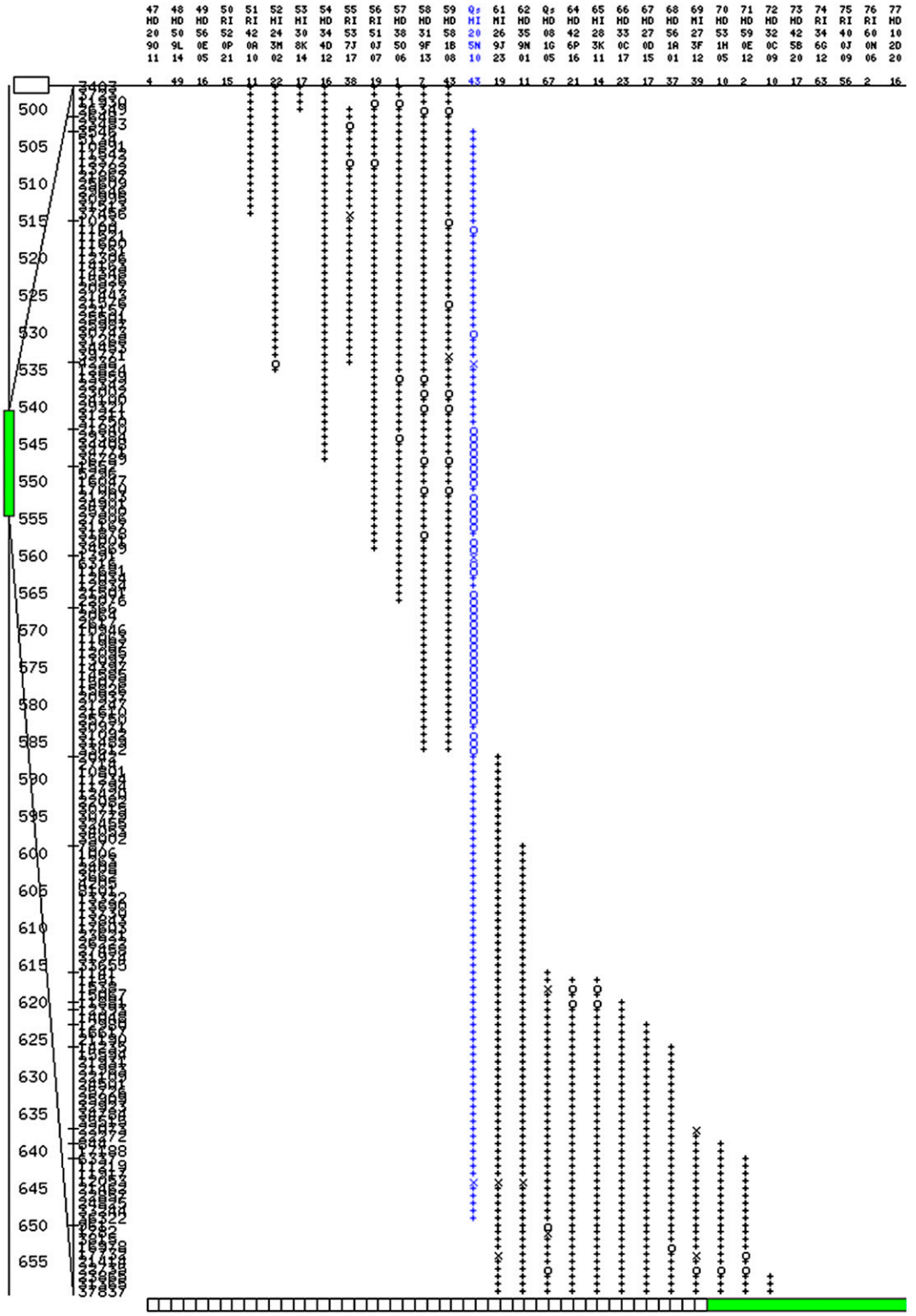

Fig. S2. A detail of the CB map for the chimeric contig ctg756 consisting of a 3D part (upper part of the contig) and 6D part (lower part of the contig), as indicated by integrated markers. BAC clones are depicted by vertical lines and restriction fragments (numbered at the left) making up BAC fingerprints are depicted by short horizontal lines across the vertical lines symbolizing BAC clones. The absence of a specific restriction fragment in a BAC fingerprint is shown by a small empty oval box on the vertical line depicting a clone. The total number of fragments not fitting clone overlaps is specified above each BAC clone. The join between the 3D and 6D portions of the contig is due to a single, questionable (Q) BAC clone (blue), which has 43 aberrant restriction fragments and missing virtually all restriction fragments near the join. FPC labeled the clone as a Q clone. The clone is likely a chimera, as evidenced by the truncation of the overlaps of BAC clone fingerprints immediately to the right and immediately to the left of the Q clone. Manual removal of the BAC clone disjoined the contig into two contigs: one located on chromosome 3D and the other located on chromosome 6D.

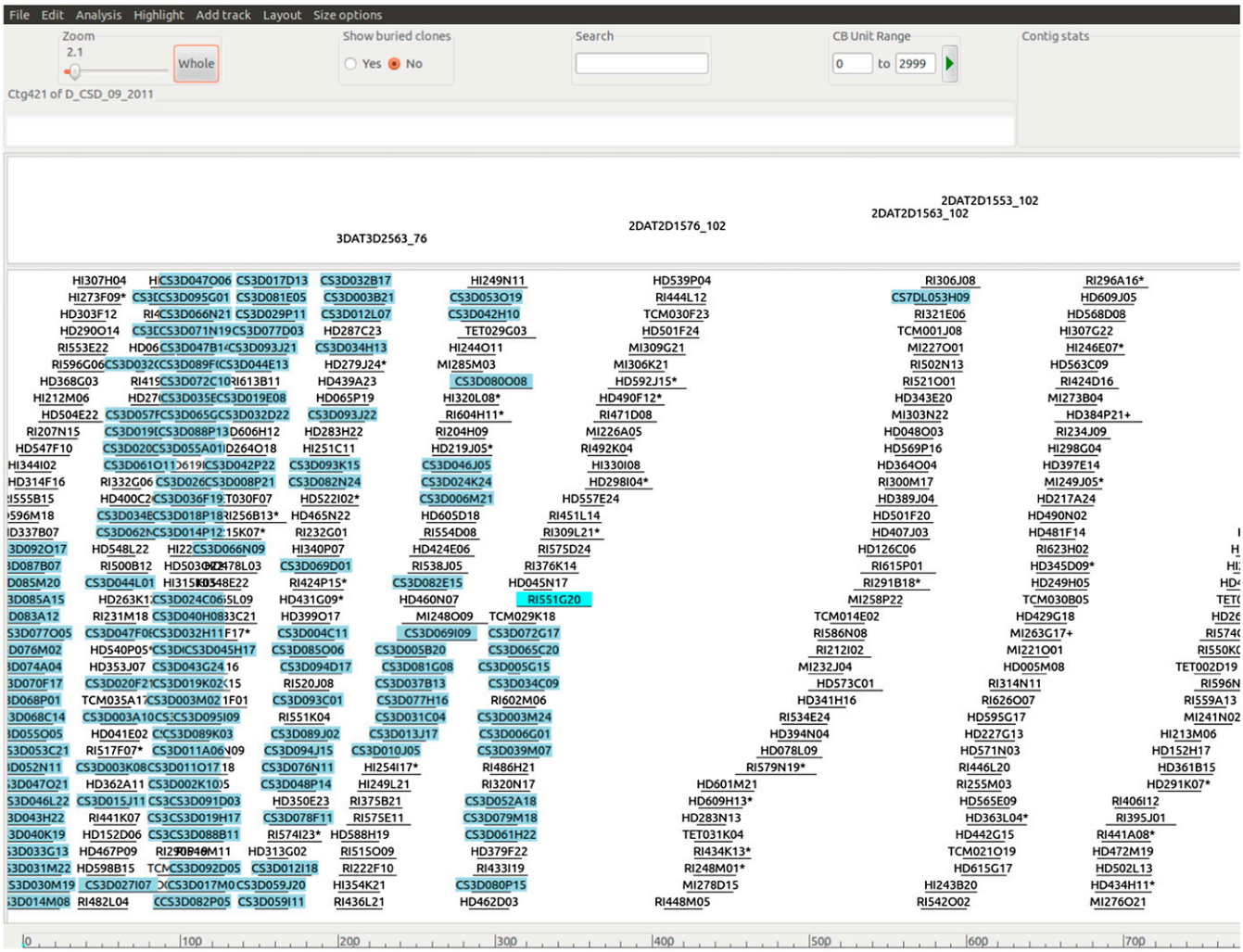

Fig. S3. *Ae. tauschii* contig ctg421 illustrating the technique of detecting chimeric contigs by coassembly of *Ae. tauschii* BAC clones with wheat D-genome BAC clones from subgenomic BAC libraries. The left side of the contig was anchored on 3DS by marker AT3D2563_76. In agreement, *Ae. tauschii* BAC clones (gray) in that area coassembled with a large number of 3DS clones (dark blue). The portion of ctg421 to the right of chimeric BAC clone RI551G20 (light blue) was devoid of 3DS BAC clones indicating that that portion of the contig came from another *Ae. tauschii* chromosome. Markers AT2D1576_102, AT2D1563_102, and AT2D1553_102 anchored that portion of the contig on *Ae. tauschii* chromosome 2D.

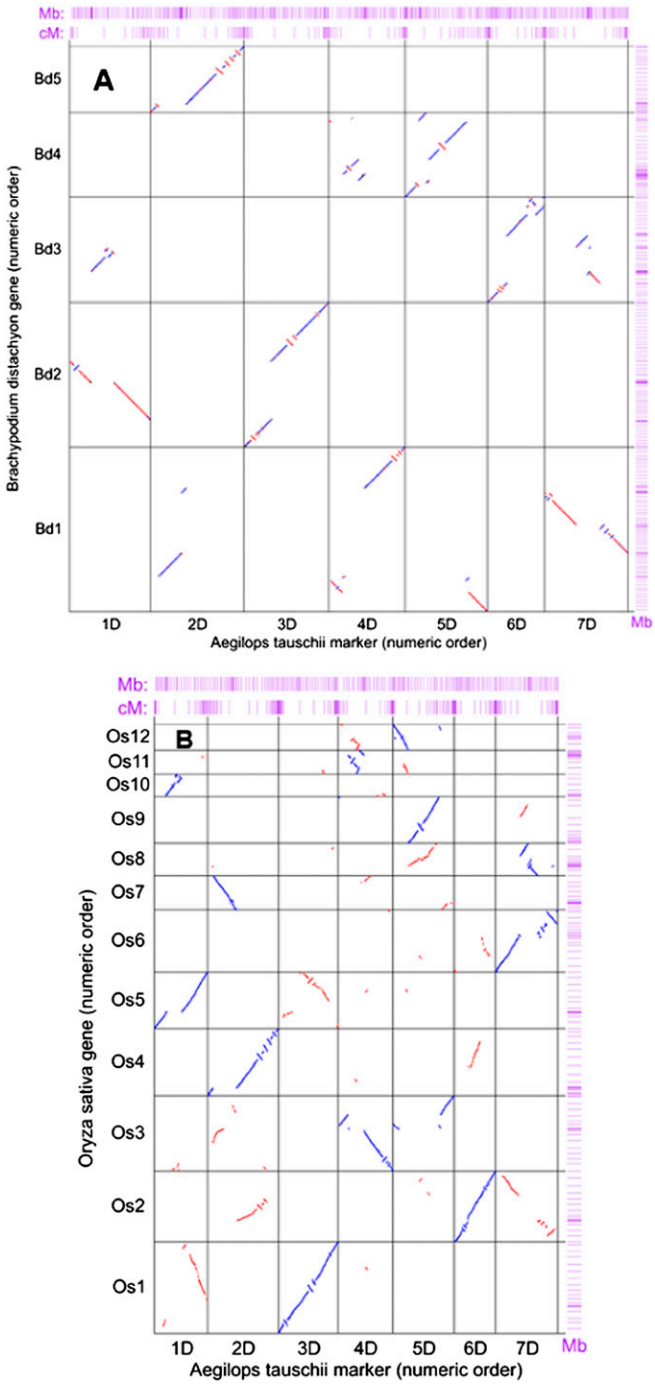

Fig. S4. *Ae. tauschii* synteny dot plots. (*A*) Pairwise comparative maps between *Ae. tauschii* and *B. distachyon*. To highlight large-scale structural rearrangements, we show only loci exhibiting collinear relationships. These loci are plotted by rank order with parallel order shown in blue and antiparallel order shown in red. Distances on the *Ae. tauschii* genetic map above the dot plots are such that each tick-mark corresponds to 10 cM. Physical distances above the dot plots refer to *Ae. tauschii,* and each tick-mark corresponds to 10 Mb. Physical distances on the right side of the dot plots refer to *B. distachyon* and each tick-mark corresponds to 1 Mb. (*B*) Synteny map between *Ae. tauschii* and rice showing both orthologous relationships (blue) and paralogous relationships (red). Syntenic paralogs are chiefly the result of the pan-grass whole genome duplication that occurred before the divergence of the rice and *Ae. tauschii* lineages, although small segmental duplications may be also present. To better highlight structural rearrangements, we show only loci exhibiting collinear relationships, which are plotted by their numeric order along chromosomes, with short arm termini at the bottom left corner. We plot distances on the *Ae. tauschii* genetic map above the dot-plots such that each tick-mark corresponds to 10 cM. We plot physical distances above and on the right side of dot-plots. For *Ae. tauschii*, each tick-mark corresponds to 10 Mb and for rice each tick-mark corresponds to 1 Mb. To recover collinear paralogs, we selected up to five significantly aligned rice genes per marker query, and omitted known orthologous relationships before mapping with DAGchainer (PMID:15247098).

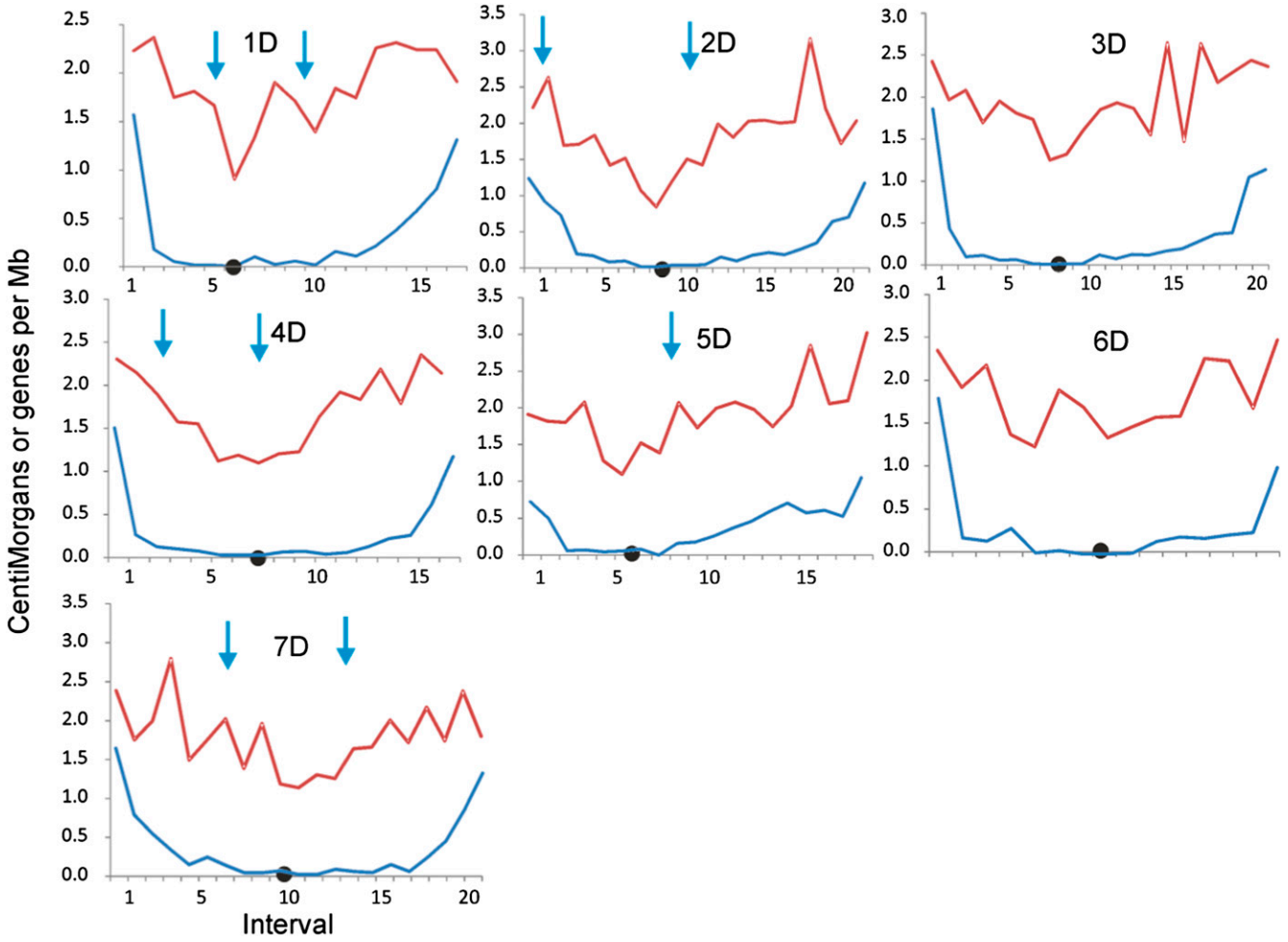

Fig. S5. Gene density expressed as the number of genes per Mb (red line) and recombination rate in centiMorgans per megabase (blue line) along the seven *Ae. tauschii* chromosomes. The short arm terminus is to the left in each graph. Black circles depict centromeres. Intervals along the *x* axis are 30 Mb long. The arrows show the sites of insertions of the ancient telomeres (chromosomes 1D, 2D, 5D, and 7D) and centromeres (4D) due to NCIs. Gene density was computed from the locations of 17,093 genes and gene fragments.

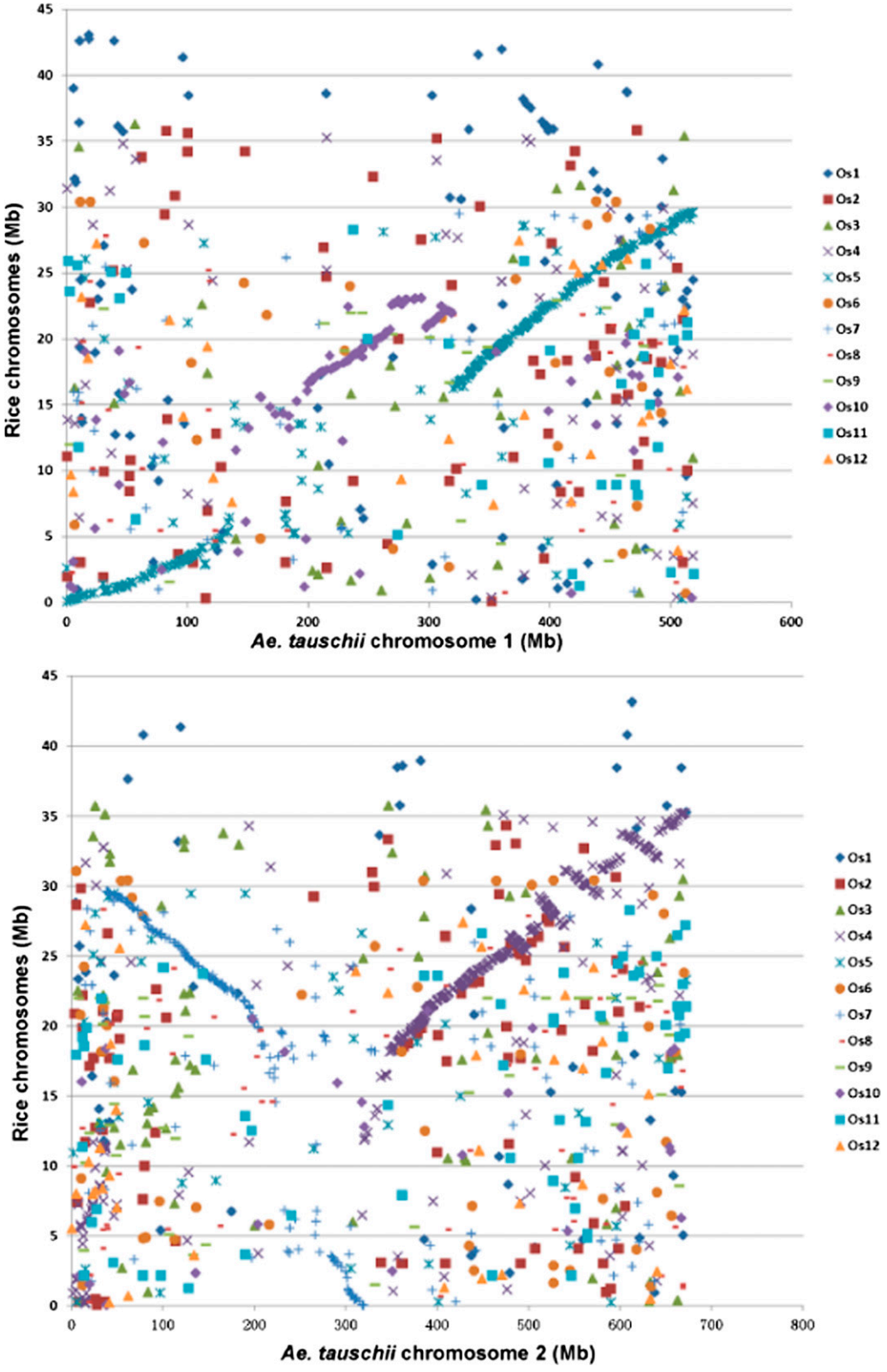

Fig. S6. (Continued)

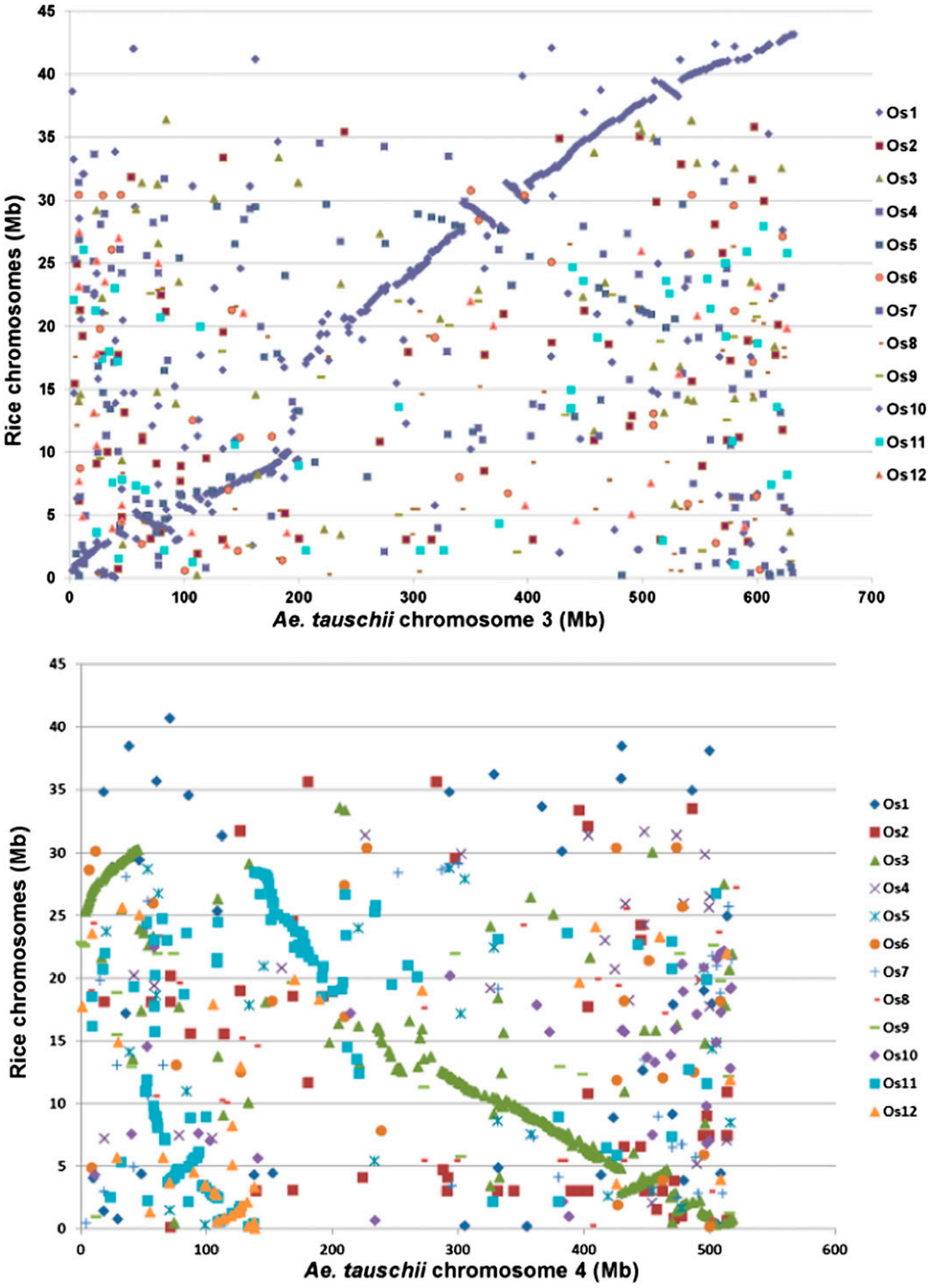

Fig. S6. (Continued)

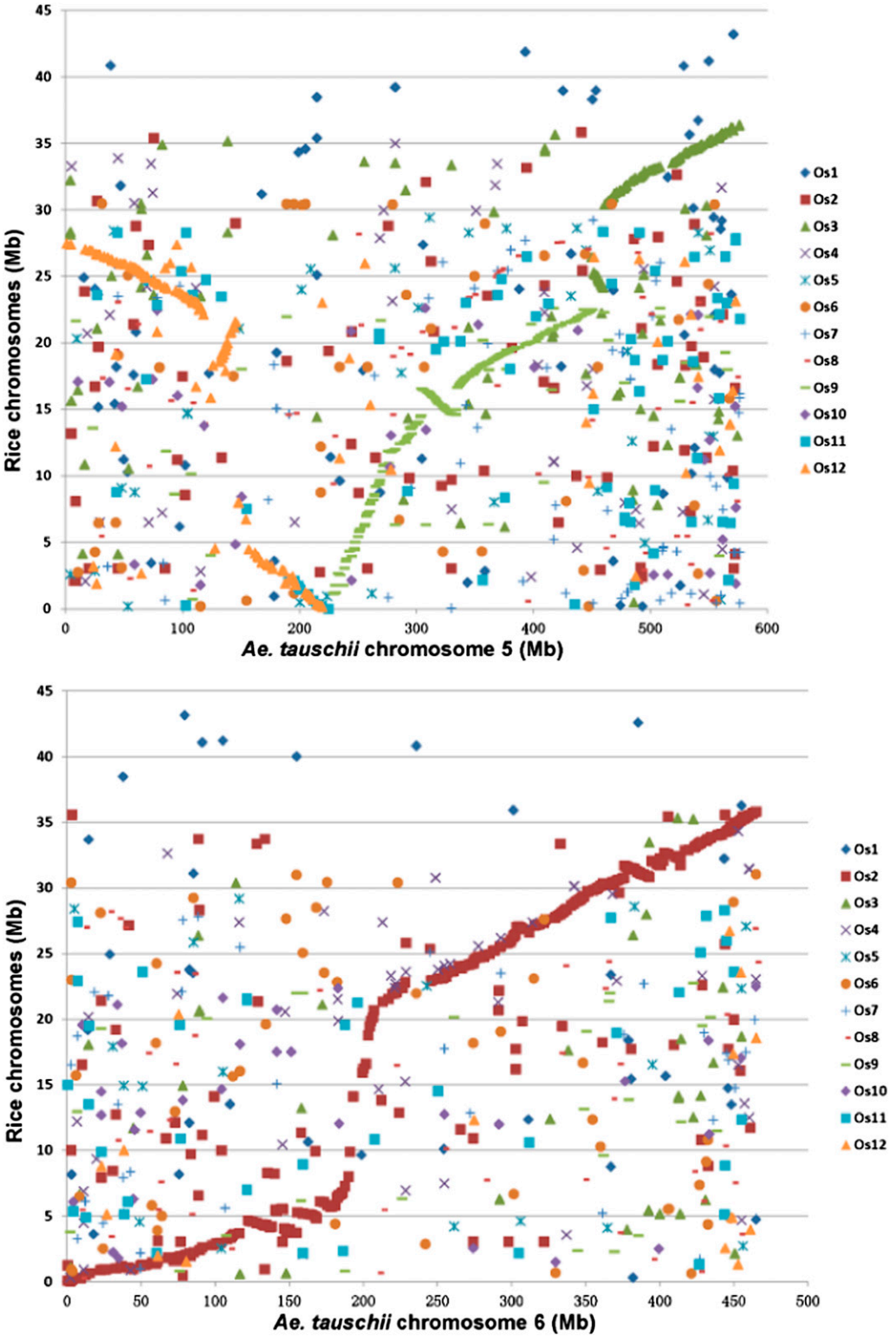

Fig. S6. (Continued)

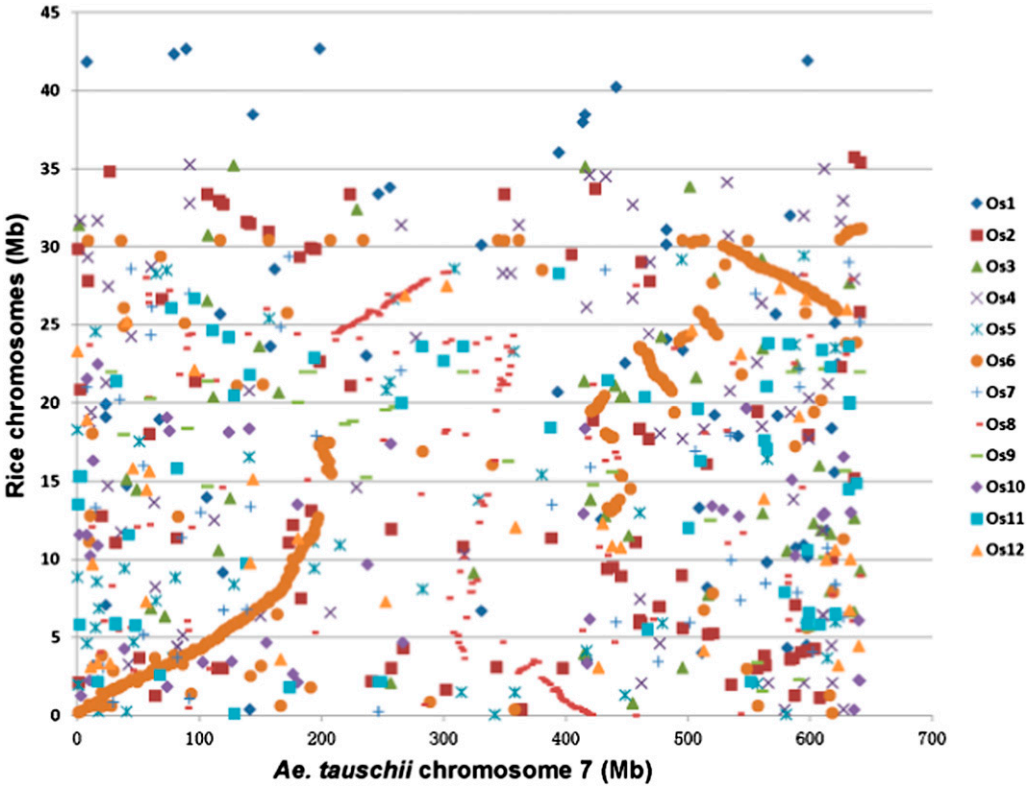

Fig. S6. Dot plots of the best matches between genes on the physical maps of the 7 *Ae. tauschii* chromosomes and 12 rice pseudomolecules. Note the concentration of noncollinear genes in the distal regions of the *Ae. tauschii* chromosomes.

Table S1. BAC and BiBAC libraries and their SNaPshot fingerprinting

Library	Name	Select. marker	Total clones	Insert size (kb)	Fingerprinted clones	Clones in contig assemblies
BamHI (phase II)	HI	Chl.	92,160	115	57,792	47,957
EcoRI (phase II)	RI	Chl.	172,800	120	155,904	132,715
HindIII (phase II)	HD	Chl.	172,800	125	158,880	143,256
MboI (phase II)	MI	Chl.	92,160	115	34,368	30,083
HindIII (phase I)	HD	Chl.		123	22,810	18,428
BAC (phase I)*	TCM	Chl.		116	20,233	17,927
BiBAC (phase I)*	TET	Tet.		114	11,572	8,935
EcoRI (phase I)	RI	Chl.		118	50	50
BamHI (phase I)	HI	Chl.		109	51	51
BamHI BiBAC (phase I)	BB	Tet.		103	12	12
HindIII BiBAC (phase I)	HB	Tet.		125	34	34
Total			529,929		461,706	399,448
Weighted insert size mean (kb)					120.5	

Chl., chloramphenicol; Tet., tetracycline.
*The library origin of the clones is unknown.

Table S2. Characterization of *Ae. tauschii* complete genes and gene fragments with respect to the presence or absence of an ortholog in at least one of the *B. distachyon*, rice, or sorghum genomes (collinear *Ae. tauschii* genes) or none (noncollinear *Ae. tauschii* genes)

Class of *Ae. tauschii* genes	Number	Percent
Total number of mapped genes and gene fragments	5,901	100.0
Collinear genes and gene fragments	3,848	65.2
Noncollinear genes	1,540	26.1
Noncollinear gene fragments	513	8.7
Total noncollinear genes and gene fragments	2,053	34.8

Table S3. Gene ontology of 4,134 mapped genes allocated to the following four groups: Collinear genes in high-recombination regions (CH), collinear genes in low-recombination regions (CL), noncollinear genes in high-recombination regions (NH), and noncollinear genes in low-recombination regions (NL)

Class	Gene ontology	CH	CL	NH	NL	CH %	CL %	NH %	NL %
GO:0000166	Nucleotide binding	124	221	66	52	10.3	11.5	10.6	13.6
GO:0003676	Nucleic acid binding	22	29	5	8	1.8	1.5	0.8	2.1
GO:0003677	DNA binding	28	50	10	14	2.3	2.6	1.6	3.7
GO:0003682	Chromatin binding	3	4	0	0	0.2	0.2	0.0	0.0
GO:0003700	Sequence-specific DNA binding transcription factor	103	151	38	27	8.6	7.8	6.1	7.1
GO:0003723	RNA binding	37	55	12	14	3.1	2.8	1.9	3.7
GO:0003774	Motor activity	11	32	2	3	0.9	1.7	0.3	0.8
GO:0003824	Catalytic activity	100	173	76	35	8.3	9.0	12.3	9.2
GO:0004518	Nuclease activity	5	16	6	5	0.4	0.8	1.0	1.3
GO:0004871	Signal transducer activity	21	45	6	4	1.7	2.3	1.0	1.0
GO:0004872	Receptor activity	27	27	19*	1	2.2	1.4	3.1	0.3
GO:0005198	Structural molecule activity	30	29	13	4	2.5	1.5	2.1	1.0
GO:0005215	Transporter activity	66	130	38	22	5.5	6.7	6.1	5.8
GO:0005488	Binding	32	53	18	9	2.7	2.7	2.9	2.4
GO:0005515	Protein binding	146	223	78	38	12.1	11.6	12.6	9.9
GO:0008135	Translation factor activity, nucleic acid binding	4	14	2	4	0.3	0.7	0.3	1.0
GO:0008289	Lipid binding	19	23	2	3	1.6	1.2	0.3	0.8
GO:0016301	Kinase activity	125	175	59	31	10.4	9.1	9.5	8.1
GO:0016740	Transferase activity	75	117	41	35	6.2	6.1	6.6	9.2
GO:0016787	Hydrolase activity	163	270	71	57	13.6	14.0	11.5	14.9
GO:0019825	Oxygen binding	16	25	19	6	1.3	1.3	3.1	1.6
GO:0030234	Enzyme regulator activity	8	9	10	1	0.7	0.5	1.6	0.3
GO:0030246	Carbohydrate binding	37	59	29	9	3.1	3.1	4.7	2.4
Total		1,202	1,930	620	382	100	100	100	100

*GO category in which the proportion of genes in the high recombination region differed at $P = 0.01$ from that in the low recombination region (Fisher exact test).