Supporting Information

Martis et al. 10.1073/pnas.1204237109

SI Materials and Methods

Purification of Chromosomes and Amplification of Their DNA. Plants with and without B chromosomes of the rye (*Secale cereale* L.) inbred line 7415 (1) were used for the analysis. Intact mitotic chromosomes were isolated by flow cytometry sorting and the purity in the sorted chromosome fractions was determined by FISH, essentially as described previously (2). Chromosomes 2R–7R of diploid rye cannot be flow-sorted individually because of the similar size of these chromosomes. The DNA of sorted chromosomes was purified and amplified by multiple displacement amplification (3).

454 Sequencing. DNA from sorted B chromosome (WCAB) and sorted chromosomes 1R–7R (WCAall), as well as genomic DNA from plants without Bs (WCAg0B) and from a plant carrying four Bs (WCAgallB), were used for 454 shotgun sequencing as described previously (4). Sequencing details are summarized in Table S1. All sequence information generated in this study is deposited in the European Nucleotide Archive (accession no. ERP001061).

Sequence Analysis. ***Repeat identification using similarity-based read clustering.*** The clustering analysis was performed as described previously (5), representing 454 read similarities of at least 90% over 55% of the read length as edges connecting the similar reads (nodes) in a virtual graph. Clusters of frequently connected nodes corresponding to groups of reads derived from different repeat families were then detected in the graph and analyzed further. To facilitate unbiased analysis, the same numbers (500,000) of randomly selected 454 reads trimmed to 200 nt were analyzed from each of the four samples (A and B chromosomes isolated by flow sorting, and rye genomic DNAs from plants containing and lacking B chromosomes). The analysis was performed simultaneously for all four samples by combining their reads before clustering. Thus, the resulting clusters included reads from different samples corresponding to the same types of repeats, allowing for straightforward comparative analysis and quantification of individual repeat types in the samples. Reads within the clusters were assembled to contigs, which were used together with cluster graph topology and sequence similarity searches for characterization of individual repeat families.

Analysis of repetitive DNA and repeat masking of sequences. Initially, the content of repetitive DNA per sequence read was identified by analysis with RepeatMasker (http://www.repeatmasker.org) against the MIPS-REdat Poaceae v8.6.1 repeat library (which contains known grass transposons from the Triticeae repeat database, http://wheat.pw.usda.gov/ITMI/Repeats, as well as de novo detected LTR retrotransposon sequences from several grass species. Subsequently, repetitive regions were masked by vmatch (http://www.vmatch.de) at the following parameters: 60% identity cutoff, 30 bp minimal length, seed length 14, exdrop 5, and e-value 0.001.

Sequence-tagged genes in the WCAB sequence dataset. To estimate the number of genes that have been captured in the WCAB and WCAall sequence collections, BLAST comparisons were carried out with the repeat-filtered reads against the *Brachypodium*, rice, and sorghum proteins/coding sequences, as well as against clustered barley EST collections (HarvEST, http://harvest.ucr.edu/; barley v1.73, assembly 35, wheat v1.19; Rice RAP-DB genome build 4, http://rapdb.dna.affrc.go.jp; sorghum genome annotation v1.4, http://genome.jgi-psf.org/Sorbi1/Sorbi1.download.ftp.html; *Brachypodium* genome annotation v1.2). The number of tagged genes and number of gene-matching reads were counted after filtering according to the following criteria: (*i*) the best hit display with a similarity greater than an adjusted species-specific similarity characteristic and (*ii*) an alignment length ≥30 amino acids (BLASTN 50 bp).

Comparative genomics to the genomic sequences of* Brachypodium*, rice, and sorghum and to the virtual genome of barley and its syntenic integration. The WCA1RS (National Center for Biotechnology Information accession no. SRX019678) (6) and WCAB datasets were compared (BLASTX) with the reference genomes of *Brachypodium*, rice, and sorghum at a filter criterion of ≥30 amino acid similarity. Matched *Brachypodium*, rice, and sorghum genes were plotted along their position on the respective chromosomes, and the average syntenic content (i.e., number of WCAB matched genes per window size of 10 genes in *Brachypodium*, rice, and sorghum) was computed and visualized in heatmaps.

Detection of A- and/or B-located SNPs. We selected sequences from the rye transcriptome dataset (7), unigene contigs available for *S. cereale* (GabiPD; accession nos. Sce_Assembly01_c1:c87199) and genic shotgun 454 reads. Consistent sequence regions from all three data sources (contigs of rye A, rye B, and rye ESTs) were identified by BLASTN. Corresponding sequences cover a total length of 1.53 Mbp and are referred as TRIPLETs. To compare the genetic distance between A and B chromosomes, reads of both datasets were mapped onto TRIPLETs using BWA (8). For A, 191.3 Mbp (886,591 reads) and 231.3 Mbp (1,112,705 reads) of B genomic reads were compared against 5,620 identified TRIPLET regions. Differences were called and analyzed using samtools (9) and vcftools (10). SNP calling identified 94,970 SNPs in A and 98,321 SNPs in B. Differences against A and B chromosome-derived sequences were seen in 6.19% and 6.41% of the TRIPLET regions, respectively. Highly stringent criteria have been used to refine variation analysis. SNPs that had a quality score of 99, a minor allele frequency >0.1, and a depth of at least 4 were considered; 30.5% of A detected SNPs and 31.5% of B detected SNPs fulfilled these criteria (Table S3). Sequences with similarity to mobile elements were excluded. To enrich for informative genic sequences, SNP-containing sequences were related to Gene Ontology (GO) assignments using BLAST and BLAST2GO. A total of 1,407 sequences have been associated to a GO term (Table S3). These were used to compare variation among A and B chromosomes. A total of 859 high-quality variations were found in A, and 1,064 were found in B; of these, 367 appeared only in A, whereas 572 were specific to B. Considering this specific diversity, our similarity study revealed that nucleotide variation in genic regions of As appear with a density of one variation per 72 bp. For genic regions in Bs, the distance between adjacent SNPs was 47 bp. In a second analysis, we identified SNP frequency in mobile elements. We used the previous strategy on 287 TRIPLET regions with annotation to transposable elements, applying identical scoring criteria (quality score >99, minor allele frequency >0.1, and depth >3). We discarded regions with collapsed read mapping that appear in high-coverage peaks (coverage >100) introduced by alignment errors of repetitive reads. We discovered similar nucleotide variation in transposable elements in A (4,714 SNPs) and B (4,352 SNPs). SNP densities were one SNP per 25 bp in A and one SNP per 26 bp in B.

Dating of origin of rye B chromosomes. The age of rye B chromosomes was determined through a SNP-based comparison between A and B chromosomes. Regions containing the 859 high-quality SNPs in A chromosomes and the final 1,064 SNPs in B chro-

mosomes were mapped onto the corresponding barley full-length cDNAs (11, 12) and the coding DNA sequence of *Brachypodium* reference genome version 1.2 using BLASTN. Regions with sequence homologies were aligned to their respective counterparts, irrespective of whether or not they had rye SNPs at the aligned positions. This procedure resulted in two datasets, one (rye EST contigs, rye A, rye B, and barley) with 16,792 alignment positions and the second (rye EST contigs, rye A, rye B, and *Brachypodium*) with 6,296 alignment positions. For both datasets, the appropriate model of sequence evolution was determined using the Akaike information criterion in Modeltest 3.7 (13), resulting in the TrN + I model of sequence evolution for the first dataset and the TrN + G model for the second. Subsequently, both datasets were phylogenetically analyzed with Bayesian inference in MrBayes 3.1.2 (14) using the appropriate models determined previously and 250,000 Markov chain Monte Carlo generations to obtain phylogenetic relationships (Fig. S3). To date, the splits of A and B genome-derived sequences in both datasets were analyzed with BEAST 1.6.1 (15) using the splits of *Brachypodium* and barley from the wheat/rye lineage at 35 My and 11.6 My, respectively, as calibration points (16). BEAST was run for each dataset twice for 100 million Markov chain Monte Carlo generations. Run parameters were checked with the Tracer program of the BEAST package, and results were compared with test convergence on the same solutions. For the BEAST analyses, the priors for ages of the calibration points were assumed to follow a lognormal distribution.

Owing to the selection process based on SNP-containing regions, the sequences used were not a totally random sample from the A and B genomes. Thus, they might contain more than the average amount of differences between As and Bs, leading to overestimates of the age of the split between these sequence groups. To control the estimated ages for A–B divergence in rye by an independent dataset, we also inferred the ages of the *Secale* species. Sequences of the ribosomal DNA internal transcribed spacer region for diploid Triticeae species plus the outgroup *Avena* were obtained from the EMBL nucleotide database, aligned, and analyzed in BEAST 1.6.1 as detailed above, using the splits between *Avena* and Triticeae (31 My; ref. 16) and the barley and wheat lineages (11.6 My) as calibration points under the GTR+G+I model of sequence evolution (Fig. S4). In this analysis, the age of the *Secale* crown group age was estimated as 1.7 My and that of the species group of *S. cereale* as ~0.8 My, slightly younger but on the same order of magnitude as the inferred split between As and Bs (1.1–1.3 My).

Analysis of organelle insertions in rye A and rye B. The repeat masked datasets of rye A and rye B reads were compared (BLASTN) against the plastid (ptDNA) and mitochondrial (mtDNA) genomes of wheat [AB042240 (ptDNA) and AP008982 (mtDNA)]. Number and identity profiles of tagged insertions were analyzed (Fig. S7). The organelle insertions were used to analyze for mechanisms involved in the insertion of ptDNA and mtDNA fragments. Sequence reads containing organelle inserts along with non–organelle-derived sequences were analyzed. A minimal length of 50 bp was required.

FISH. Mitotic chromosome spreads of rye were prepared and in situ hybridization with the selected probes was performed as described previously (17).

1. Jimenez MM, Romera F, Puertas MJ, Jones RN (1994) B chromosomes in inbred lines of rye (*Secale cereale* L), 1: Vigor and fertility. *Genetica* 92:149–154.
2. Kubaláková M, et al. (2003) Analysis and sorting of rye (*Secale cereale* L.) chromosomes using flow cytometry. *Genome* 46:893–905.
3. Simková H, et al. (2008) Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. *BMC Genomics* 9:294.
4. Mayer KFX, et al. (2009) Gene content and virtual gene order of barley chromosome 1H. *Plant Physiol* 151:496–505.
5. Novák P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* 11:378.
6. Fluch S, et al. (2012) Sequence composition and gene content of the short arm of rye (*Secale cereale*) chromosome 1. *PLoS ONE* 7:e30784.
7. Haseneyer G, et al. (2011) From RNA-seq to large-scale genotyping–genomics resources for rye (*Secale cereale* L.). *BMC Plant Biol* 11:131.
8. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595.
9. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25:2078–2079.
10. Danecek P, et al.; 1000 Genome Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
11. Matsumoto T, et al. (2011) Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries. *Plant Physiol* 156:20–28.
12. Sato K, et al. (2009) Development of 5006 full-length CDNAs in barley: A tool for accessing cereal genomics resources. *DNA Res* 16:81–89.
13. Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.
14. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
15. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
16. Chalupska D, et al. (2008) Acc homoeoloci and the evolution of wheat genomes. *Proc Natl Acad Sci USA* 105:9691–9696.
17. Ma L, et al. (2010) Synteny between *Brachypodium distachyon* and *Hordeum vulgare* as revealed by FISH. *Chromosome Res* 18:841–850.

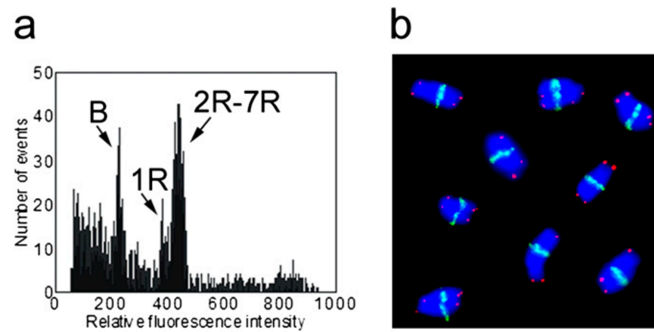

Fig. S1. Flow sorting of rye B chromosomes. (*A*) Histogram of fluorescence intensity (flow karyotype) obtained after flow cytometry analysis of DAPI-stained mitotic metaphase chromosomes isolated from root tips of seedlings with Bs. The peak representing Bs is well discriminated from the remaining chromosomes (1R–7R), which form a composite peak with a shoulder of 1R. (*B*) Examples of sorted Bs after fluorescent labeling of *Afa* repeats (yellow-green) and *Arabidopsis*-type telomere repeat (red) using FISH. The sorted fractions consisted mainly of Bs (mean, 95.5% ± 0.7%) as determined by FISH on 1,000 sorted chromosomes obtained during each sort run. The contamination was related to various chromosomes and chromosome fragments.

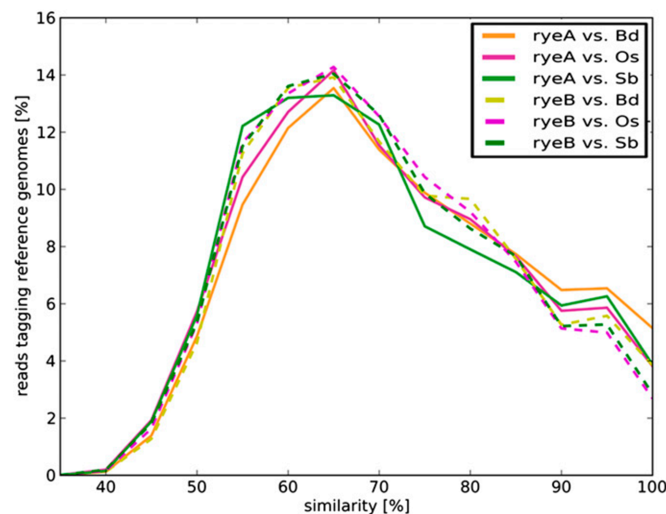

Fig. S2. Sequence comparison of rye A and B reads against rice (Os), *Brachypodium distachion* (Bd), and sorghum (Sb) genes. Rye A and B sequences have been compared against the complete genome complements of rice, sorghum, and *B. distachion*, respectively. Similarity comparisons were undertaken using BLASTX searches. The respective best-matching homolog with an alignment length of at least 30 amino acids was selected. The observed identities were grouped into five bins and plotted on the *x*-axis.

Fig. S3. Phylogenetic relationships and age estimates for rye A and B chromosomes. Sequences derived from A and B chromosomes of rye were aligned with homologus parts derived from *S. cereale* EST data, *Hordeum vulgare* EST data, and *B. distachyon* genomic sequences. Bayesian phylogenetic analysis was conducted in BEAST 1.6.1 with nodes fixed at 35 million years (My) for the split between *Brachypodium* and the other taxa and at 11.6 MY for the split between *Hordeum* and *Secale*. Numbers along branches depict Bayesian posterior probabilities derived from phylogenetic analyses in MrBayes, the bold numbers at nodes are ages (in MY) estimated by BEAST, bars give the 95% highest posterior density (HPD) assuming lognormally distributed age priors, and red dots indicate the calibration points. (*A*) Tree derived from a dataset of sequences from all groups. (*B*) Tree derived from a dataset of sequences from *Hordeum* and *Secale*.

Avena longiglumis
Psathyrostachys juncea
Ps. fragilis
Ps. huashanica
Hordeum bulbosum
H. bulbosum
H. vulgare subsp. vulgare
H. vulgare subsp. spontaneum
H. murinum subsp. glaucum
H. marinum
H. gussoneanum
H. marinum
H. gussoneanum
H. californicum
H. californicum
H. brevisubulatum subsp. violaceum
H. bogdanii
H. roshevitzii
Heteranthelium piliferum
Dasypyrum villosum
Peridictyon sanctum
Pseudoroegneria spicata
Pseud. libanotica
Australopyrum pectinatum
Aus. pectinatum
Agropyron mongolicum
Agr. puberulum
Agr. cristatum
Taeniatherum caput-medusae
Secale sylvestre (AJ409211)
S. sylvestre (AJ409210)
S. strictum subsp. kuprijanovii (AJ409205)
S. strictum subsp. anatolicum (AJ409208)
S. strictum subsp. africanum (AJ409209)
S. vavilovii (AJ608152)
S. vavilovii (AJ409204)
S. cereale (AF303400)
S. cereale (L36504)
S. strictum subsp. africanum (L36503)
S. cereale subsp. ancestrale (AJ409199)
S. cereale subsp. cereale (AJ409202)
Thinopyrum bessarabicum
Lophopyrum elongatum
Crithopsis delileana
Triticum urartu
T. urartu
T. monococcum
T. baeoticum
T. baeoticum
T. baeoticum
T. baeoticum
T. baeoticum
T. baeoticum
Eremopyrum bonaepartis
Henrardia persica
Aegilops searsii
Ae. sharonensis
Ae. speltoides
Ae. speltoides
Ae. longissima
Ae. bicornis
Ae. tauschii
Ae. speltoides
Ae. uniaristata
Ae. mutica
Ae. comosa
Ae. markgrafii
Ae. comosa
Ae. umbellulata
1.7
0.76
30 25 20 15 10 5 0
million years before present

Fig. S4. Phylogenetic tree (chronogram) of Triticeae grasses based on sequences of the nuclear ribosomal DNA internal transcribed spacer region. Phylogenetic relationships and age estimations (in My) were calculated with BEAST 1.6.1. The splits between *Avena* and Triticeae (31 My) and *Hordeum* and *Triticum* (11.6 My) were used as calibration points (marked by red dots). Blue bars at nodes indicate the 95% HPD distribution. The analysis yielded a crown age of *Secale* of 1.7 My (0.6–3.0 My 95% HPD) and an age for the split between *Secale sylvestre* and the other taxa within *Secale* of 0.76 My (0.25–1.37 My 95% HPD).

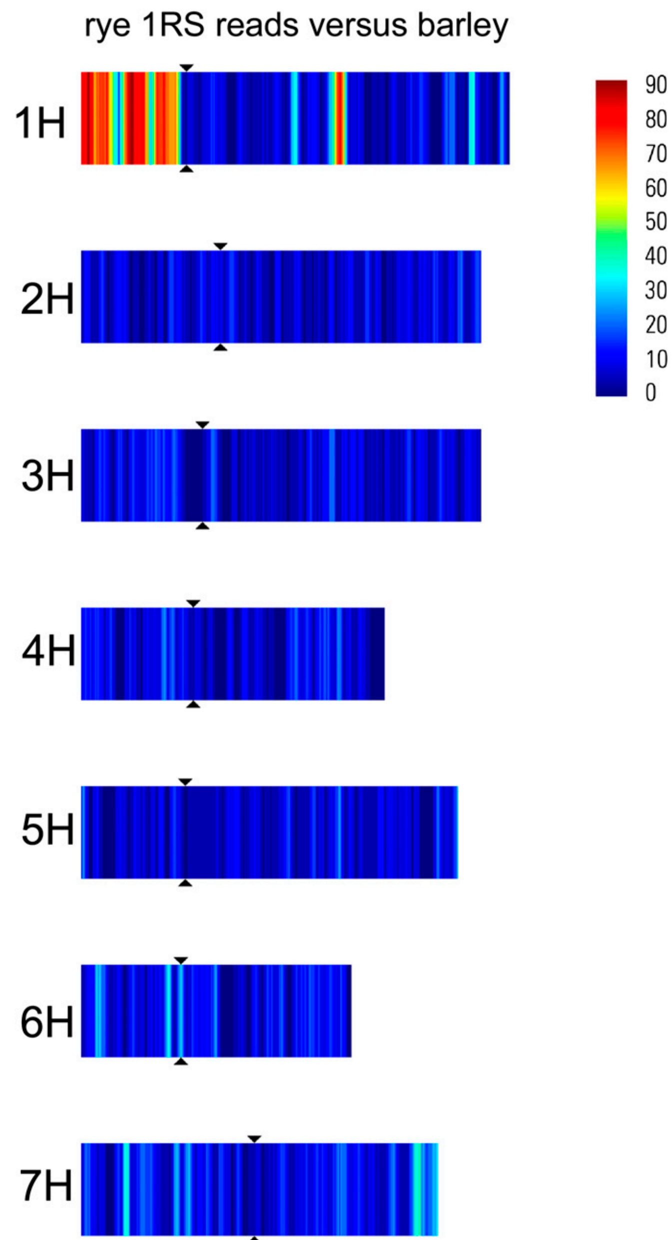

Fig. S5. Rye 1RS sequence reads (6) mapped on the barley genome. The heatmap depicts the location of detected barley syntenic regions. Sequence reads were anchored on barley using BLASTX and the best detectable match. Individual chromosomes were numbered. Barley chromosome 1H is the region with conserved gene content to rye chromosome 1RS. A triangle indicates the centromere position on each barley chromosome.

Fig. S6. Comparison of proportions of the most abundant repetitive elements between flow-sorted B and A chromosomes and between genomic DNAs from individuals carrying (+B) or lacking (−B) B chromosomes, respectively. Each horizontal bar represents one repeat cluster, and the dark/light shading shows the ratio of cluster sizes in the compared samples. Clusters are sorted according to repeat types and lineages, and their cumulative proportions in the analyzed samples are shown in the right panel. Cluster order corresponds to the data provided in Table S4.

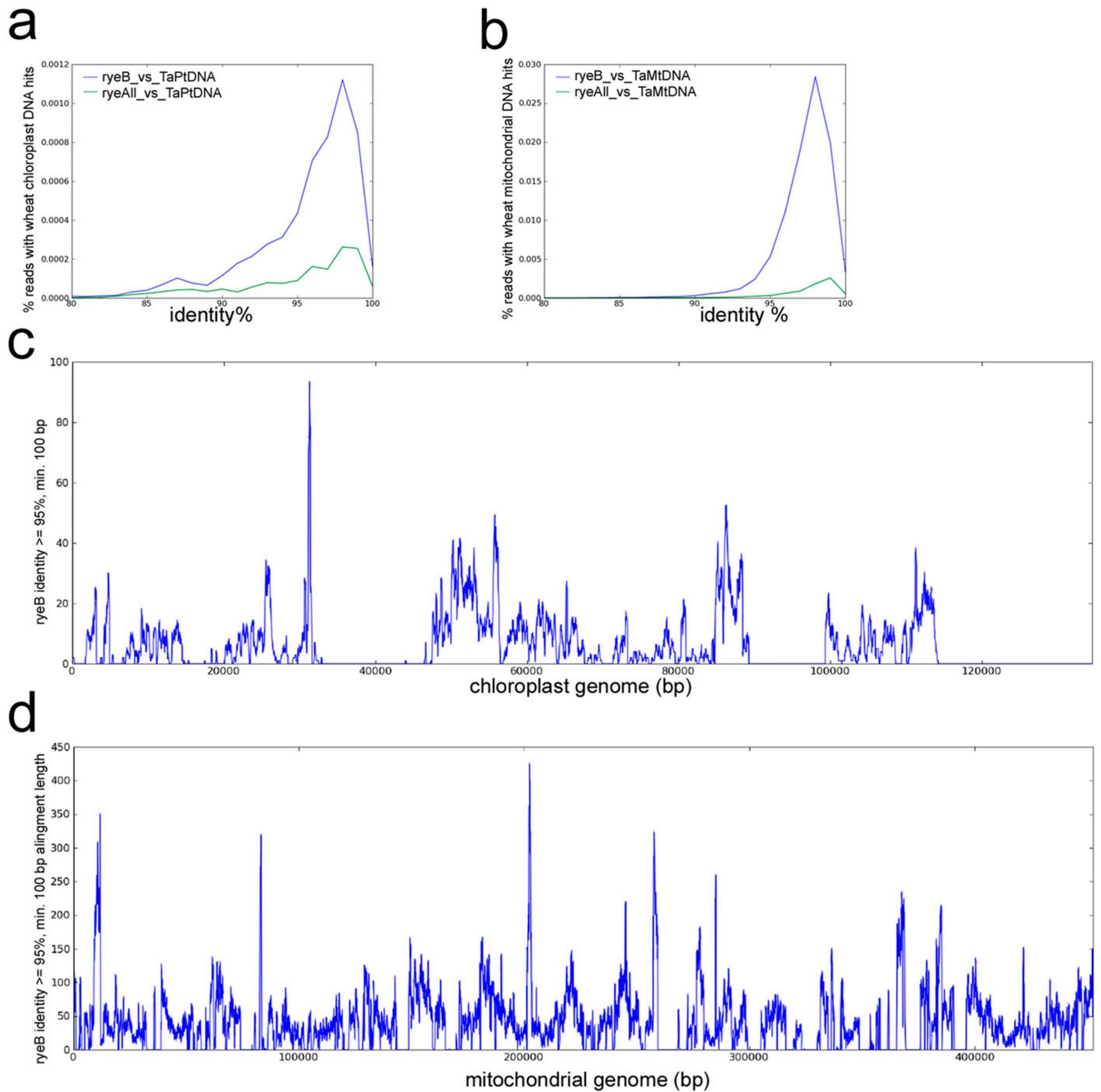

Fig. S7. Comparison of rye A and rye B reads against organellar genomes. The normalized results of comparisons of repeat-masked rye A and rye B against the complete genomes of wheat organelles are shown. (*A* and *B*) Comparisons with the wheat chloroplast genome are indicated by TaPt (*A*), and those with the wheat mitochondrial genome are indicated by TaMt (*B*). The observed identity profiles indicates that rye B carries a higher number of chloroplast and mitochondrial DNA insertions compared with rye A. (*C* and *D*) Comparisons of 100-bp-long rye B reads with the wheat chloroplast (*C*) and the wheat mitochondrial genome (*D*). Almost all parts of the organellar genome are shared by the B chromosome.

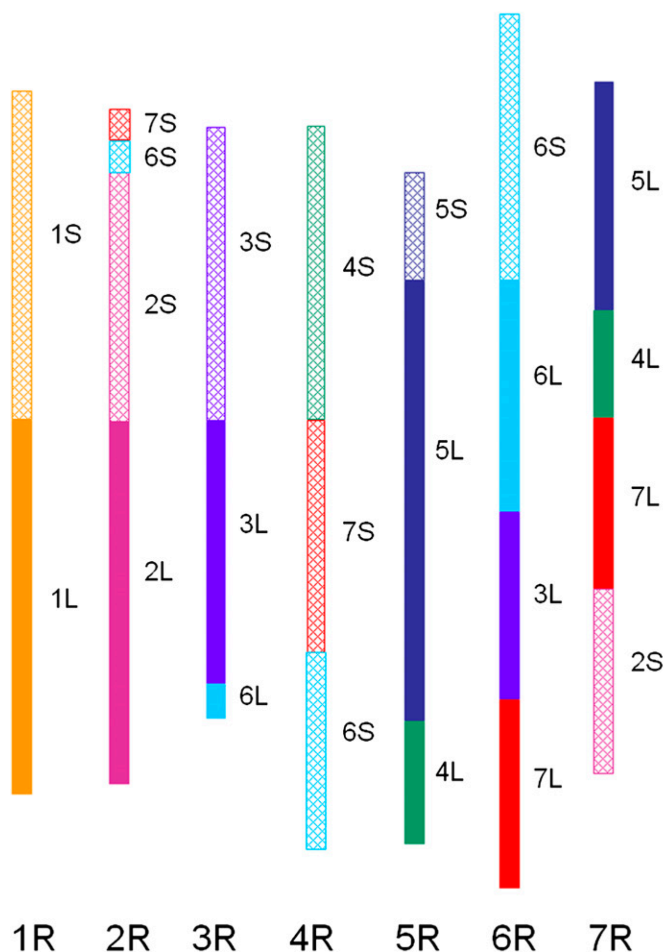

Fig. S8. *S. cereale* chromosomes and their relationships with the homeologous chromosomes of wheat and barley modified (1–3).

1. Devos KM, et al. (1993) Chromosomal rearrangements in the rye genome relative to that of wheat. *Theor Appl Genet* 85:673–680.
2. Mayer KF, et al. (2011) Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell* 23:1249–1263.
3. Hernandez P, et al. (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.

Table S1. 454 sequence read characteristics

Dataset	Total base pairs		Median read length		Coverage		Lander-Waterman expectation of detected bp		% HQ sequences after repeat masking
	All	HQ	All	HQ	All	HQ	All	HQ	
Sorted rye B (WCAB)	838,660,944	521,476,205	552	343	1.44×	0.9×	76.3%	59.3%	39.6%
Sorted rye all As (WCAall)	532,362,715	379,952,169	527	377	0.07×	0.05×	6.8%	4.8%	32.4%
gDNA 0B (WCAg0B)	594,460,439	416,597,333	523	367	0.08×	0.05×	7.7%	4.8%	27.5%
gDNA 4B (WCAgallB)	914,808,184	663,341,461	527	382	0.09×	0.06×	8.5%	6.3%	27.5%

Summary of the sequence read characteristics obtained by 454 sequencing of sorted rye chromosome B exclusively (WCAB), sorted pooled rye chromosome As (WCAall), genomic rye 0B DNA (WCAg0B), and genomic rye +4B DNA (WCAgallB). HQ, high-quality sequences after soft clipping of primer and low-quality sequence. The reference size is 580 Mbp for rye B and 7.92 Gbp for the complete rye genome (rye A).

Table S2. Comparison of ryeA/ryeB sequences against three model genome resources

	Coverage	No. reads with hits	% reads with hits	No. reads with hits (≥30 aa/100 bp)	No. reads with hits (≥70% and ≥30 aa/100 bp)	No. nonredundant genes (≥70% and ≥30 aa/100 bp)
Rye A						
Bd	0.05×	33,292	10.2	31,389	13,773	5,981
Os		29,530	9	28,041	10,921	4,957
Sb		30,378	9.3	28,872	10,957	5,274
Rye B						
Bd	0.9×	48,467	8.1	45,861	18,268	4,189
Os		42,136	7	40,161	14,661	3,449
Sb		42,708	7	40,724	14,821	3,815

Nucleotide and protein sequence directed comparisons between ryeA/ryeB and the *Brachypodium* (Bd), rice (Os), and sorghum (Sb) genes were used to analyze the genes detected by the two datasets.

Table S3. Summary of SNP filters

	SNPs		INDELs		Contigs (unigenes)	
Features	A	B	A	B	A	B
	94,970	98,321	7,715	7,675	3,101	3,185
SNP quality (99)	57,980	62,858	6,110	6,184	2,348	2,541
Minor allele frequency (>0.1)	57,944	62,802	6,110	6,184	2,348	2,541
Read coverage (4–100)	28,943	30,986	3,580	3,398	2,222	2,416
Feature-based selection	859	1,064	122	119	81	84

The table shows the reduction of detected SNP candidates based on the mapping of genomic A and B reads against rye transcriptome contigs published in GabiPD. Putative SNP candidates fulfill (*i*) a SNP quality score of 99, (*ii*) a minor allele frequency (MAF) of >0.1, and (*iii*) a read coverage between 4–100. (*iv*) A removal of sequences with BLAST hits against mobile elements led to the final set of SNPs, where regions have adequate GO-term assignment. A, all A chromosomes; B, B chromosome.

Table S4. Proportion of major repeat families in analyzed samples calculated from clustering analysis of 454 reads

Cluster	Repeat type	Family	B chromosomes, %	A chromosomes, %	Genomic +Bs, %	Genomic −Bs, %	B/A ratio
5	Ty3/gypsy	Athila	6.12	8.23	4.11	4.39	0.74
1	Ty3/gypsy	Athila	10.30	14.97	6.10	6.83	0.69
2	Ty3/gypsy	Athila	5.66	10.99	6.78	7.42	0.52
59	Ty3/gypsy	Athila	0.08	0.17	0.08	0.08	0.47
9	Ty3/gypsy	Chromo	2.13	1.26	2.54	2.17	1.69
62	Ty3/gypsy	Chromo	0.05	0.03	0.17	0.14	1.67
11x	Ty3/gypsy	Chromo	1.58	1.09	1.85	1.88	1.45
3	Ty3/gypsy	Chromo	6.99	6.04	8.39	7.63	1.16
20	Ty3/gypsy	Chromo	0.31	0.32	1.05	1.11	0.97
12	Ty3/gypsy	Chromo	0.85	0.96	2.62	2.76	0.89
18	Ty3/gypsy	Chromo	0.48	0.64	0.98	1.02	0.75
75	Ty3/gypsy	Chromo	0.03	0.05	0.07	0.09	0.60
41	Ty3/gypsy	Chromo	0.16	0.28	0.28	0.29	0.57
8	Ty3/gypsy	Chromo CR	5.13	2.83	1.39	1.09	1.81
6	Ty3/gypsy	Chromo CR	4.17	2.73	2.88	2.40	1.53
38	Ty3/gypsy	Tat/Ogre	0.37	0.27	0.24	0.20	1.37
10	Ty3/gypsy	Tat/Ogre	2.20	2.12	2.78	2.65	1.04
35	Ty3/gypsy	Tat/Ogre	0.16	0.16	0.44	0.44	1.00
66	Ty3/gypsy	Tat/Ogre	0.09	0.12	0.05	0.05	0.75
7	Ty3/gypsy	Tat/Ogre	0.55	0.74	5.05	4.60	0.74
50	Ty3/gypsy	Tat/Ogre	0.05	0.07	0.26	0.22	0.71
46	Ty3/gypsy	Tat/Ogre	0.13	0.19	0.24	0.27	0.68
42	Ty3/gypsy	Tat/Ogre	0.12	0.19	0.33	0.34	0.63
44	Ty3/gypsy	Tat/Ogre	0.06	0.12	0.36	0.38	0.50
74	Ty3/gypsy	Tat/Ogre	0.01	0.03	0.11	0.11	0.33
73	Ty3/gypsy		0.06	0.03	0.10	0.09	2.00
82	Ty3/gypsy		0.01	0.01	0.08	0.08	1.00
43	Ty3/gypsy		0.14	0.22	0.28	0.32	0.64
92	Ty3/gypsy		0.03	0.05	0.03	0.04	0.60
29	Ty3/gypsy		0.32	0.54	0.34	0.36	0.59
67	Ty3/gypsy		0.07	0.12	0.06	0.07	0.58
60	Ty3/gypsy		0.04	0.09	0.13	0.14	0.44
48	Ty3/gypsy		0.11	0.25	0.16	0.20	0.44
4	Ty1/copia	Angela	3.05	2.59	9.00	9.06	1.18
70	Ty1/copia	Angela	0.02	0.03	0.12	0.12	0.67
15	Ty1/copia	Maximus/SIRE	1.05	1.12	1.18	1.10	0.94
16	Ty1/copia	Maximus/SIRE	0.72	0.92	1.13	1.22	0.78
13	Ty1/copia	Maximus/SIRE	1.02	1.49	2.10	2.21	0.68
11	Ty1/copia	Bianka	2.40	0.03	0.40	0.01	80.00
86	Ty1/copia	Bianka	0.04	0.02	0.06	0.05	2.00
36	Ty1/copia	TAR	0.36	0.12	0.39	0.26	3.00
53	Ty1/copia	Tork	0.33	0.08	0.08	0.04	4.13
89	Ty1/copia		0.06	0.03	0.04	0.03	2.00
95	Ty1/copia		0.02	0.03	0.05	0.04	0.67
19	DNA transposon		0.92	0.56	0.68	0.65	1.64
49	DNA transposon		0.17	0.11	0.17	0.15	1.55
33	DNA transposon		0.26	0.21	0.42	0.44	1.24
37	DNA transposon		0.22	0.20	0.36	0.35	1.10
32	DNA transposon		0.29	0.27	0.43	0.43	1.07
78	DNA transposon		0.04	0.06	0.06	0.07	0.67
56	DNA transposon		0.08	0.13	0.12	0.13	0.62
24	DNA transposon		0.43	0.70	0.43	0.50	0.61
17	DNA transposon		0.72	1.19	0.70	0.76	0.61
72	DNA transposon		0.06	0.11	0.05	0.06	0.55
31	DNA transposon		0.27	0.56	0.28	0.32	0.48
40	DNA transposon		0.22	0.46	0.18	0.21	0.48
52	DNA transposon		0.09	0.20	0.12	0.14	0.45
63	Satellite		0.31	0.00	0.07	0.00	—
21	Satellite	E3900	2.18	0.03	0.36	0.00	72.67
26	Satellite		1.32	0.03	0.44	0.00	44.00
9	Satellite	D1100	1.51	0.09	0.52	0.07	16.78
93	Satellite		0.05	0.02	0.05	0.04	2.50

Table S4. Cont.

Cluster	Repeat type	Family	B chromosomes, %	A chromosomes, %	Genomic +Bs, %	Genomic −Bs, %	B/A ratio
76	Satellite		0.04	0.12	0.04	0.04	0.33
23	Satellite	pSC250	0.06	0.18	0.83	1.11	0.33
27	Satellite	pSC200	0.07	0.23	0.35	1.01	0.30
68	Satellite	pSc119	0.01	0.10	0.03	0.17	0.10
47	Satellite	Tail	0.01	0.13	0.30	0.33	0.08
28+54+65+80+81+88 +91+95+39part	Organellar	Mitochondria	2.65	0.14	0.47	0.22	18.93
14+39part	Organellar	Plastid	1.35	0.11	1.74	2.72	12.27
83	Other/unclassified		0.15	0.00	0.03	0.00	—
97	Other/unclassified		0.11	0.00	0.02	0.00	—
55	Other/unclassified		0.42	0.01	0.05	0.00	42.00
64	Other/unclassified		0.16	0.06	0.08	0.07	2.67
87	Other/unclassified		0.08	0.04	0.02	0.02	2.00
90	Other/unclassified		0.08	0.04	0.02	0.02	2.00
69	Other/unclassified		0.10	0.06	0.08	0.07	1.67
84	Other/unclassified		0.04	0.03	0.06	0.05	1.33
22	Other/unclassified		0.82	0.63	0.54	0.52	1.30
57	Other/unclassified		0.14	0.13	0.09	0.09	1.08
51	Other/unclassified		0.10	0.10	0.17	0.18	1.00
96	Other/unclassified		0.04	0.05	0.02	0.02	0.80
85	Other/unclassified		0.05	0.07	0.02	0.03	0.71
25	Other/unclassified		0.45	0.66	0.42	0.44	0.68
61	Other/unclassified		0.10	0.15	0.07	0.08	0.67
30	Other/unclassified		0.32	0.60	0.29	0.32	0.53
58	Other/unclassified		0.08	0.15	0.10	0.12	0.53
77	Other/unclassified		0.04	0.08	0.06	0.07	0.50
34	Other/unclassified		0.23	0.50	0.25	0.27	0.46
79	Other/unclassified		0.03	0.09	0.04	0.04	0.33
45	Other/unclassified.		0.01	0.07	0.38	0.40	0.14

This table provides proportions (%) of reads from each sample (100% = 500,000) assigned to individual clusters and classification of repetitive sequences in the clusters. This information is provided for the 90 largest clusters. (Mitochondrial and plastid sequences are each represented by one cluster only, generated by merging several smaller clusters.) Clusters are sorted based on repeat type and proportion in B chromosomes compared with A chromosomes; the order of clusters is the same as in Fig. S6. —, not applicable.