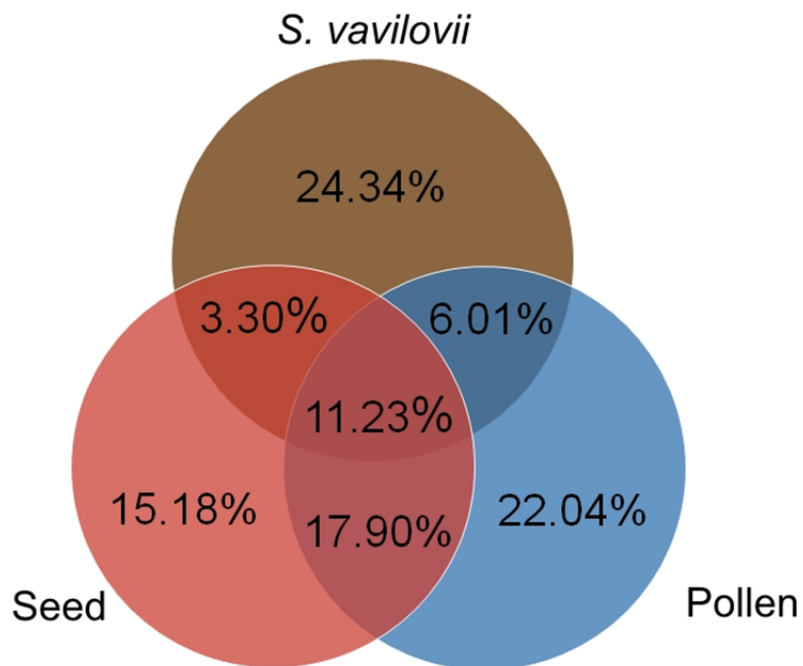
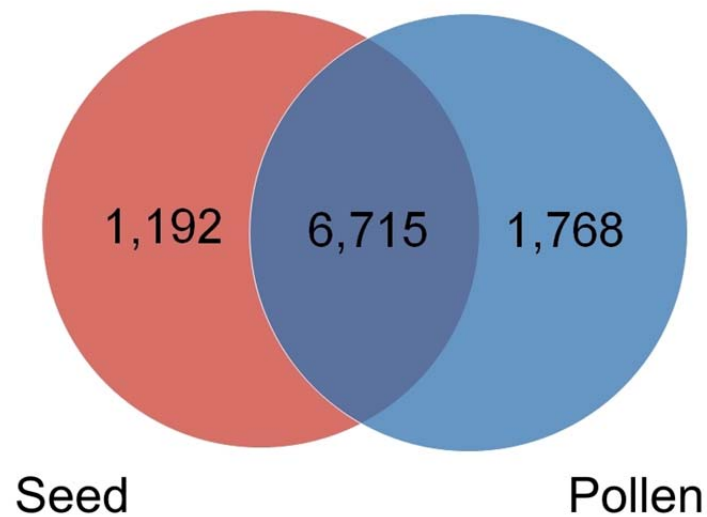


## Supplementary figures



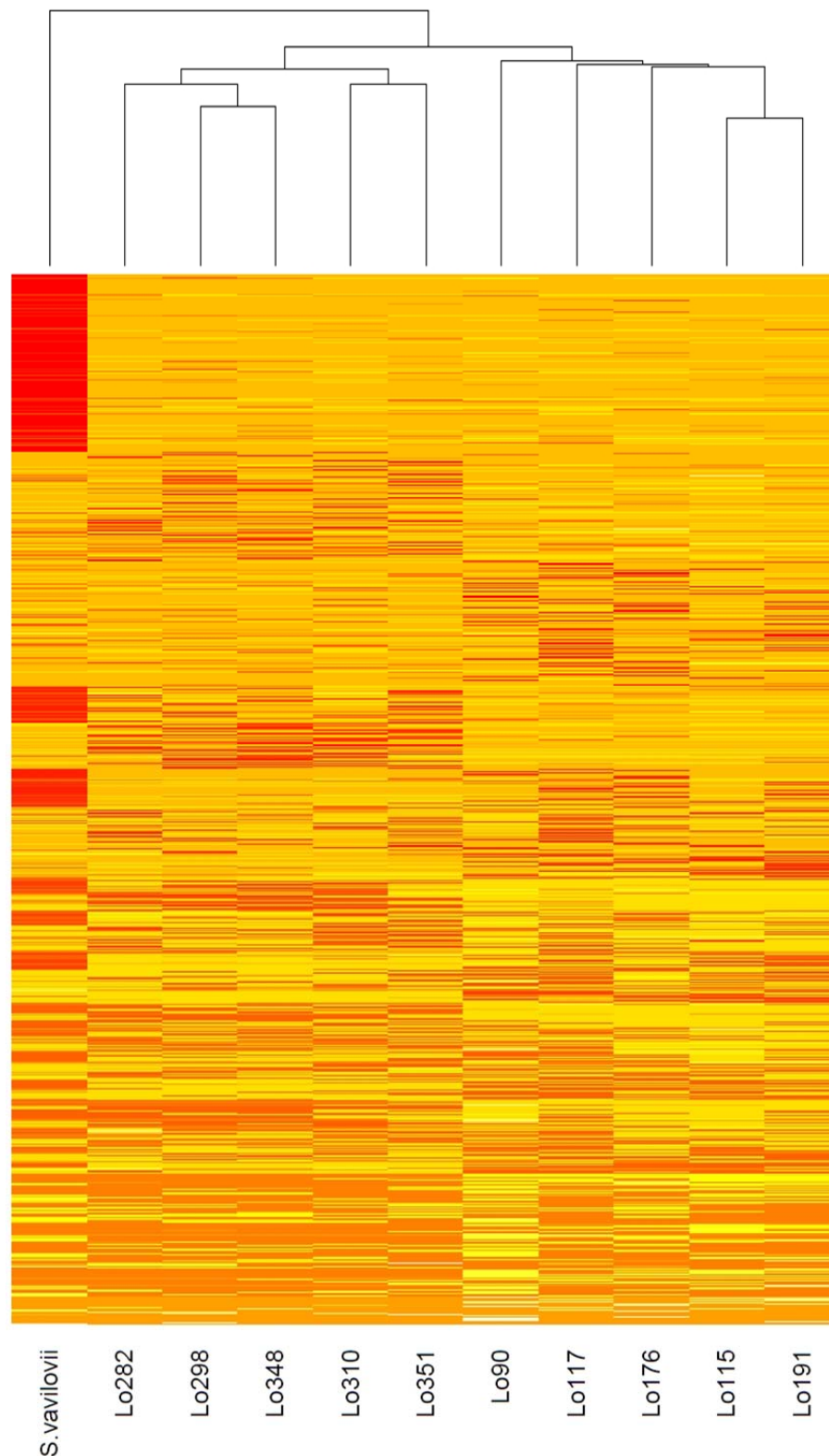
**Figure S1 Comparisons of diversity between two different breeding pools and *S. vavilovii***

For the total number of SNVs (8,626,622) the percentage of shared and pool- or species-specific polymorphisms is shown for seed (red) and pollen (blue) parent pool and *S. vavilovii* (brown).



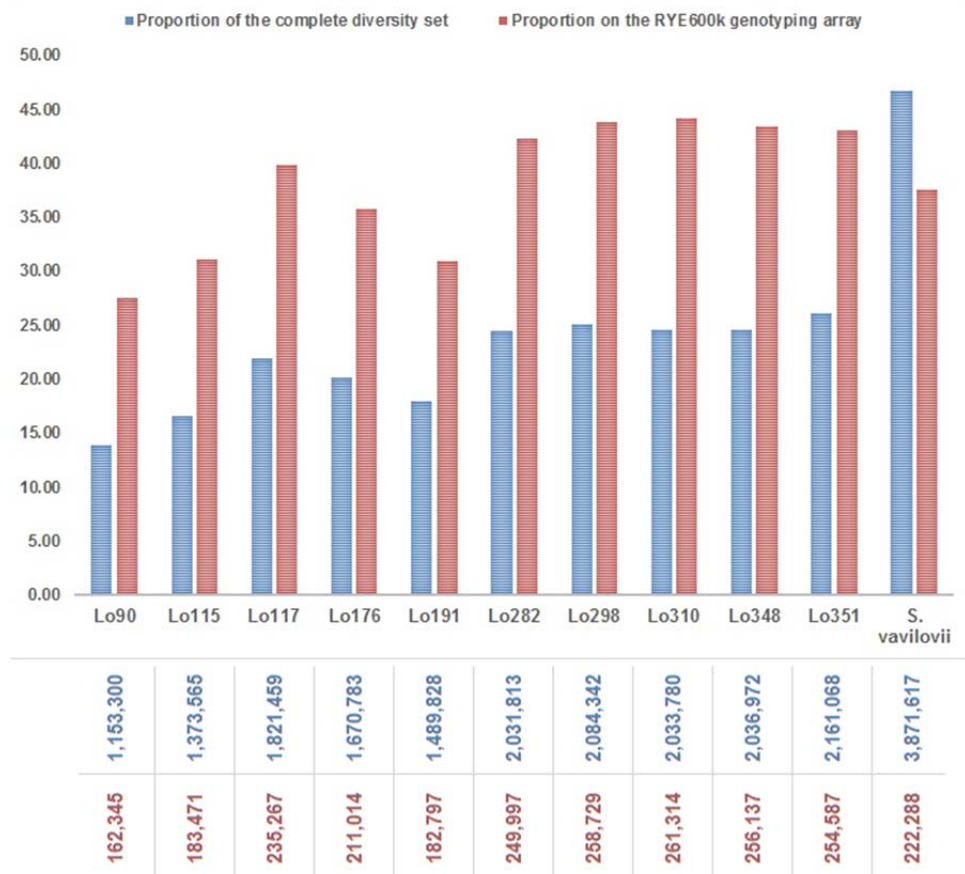
**Figure S2 Comparison of rye genes with non-synonymous SNVs (nsSNVs) in the seed and pollen parent pools**

The Venn diagram shows the number of genes that are affected by nsSNV mutation(s) in at least one rye inbred line of the seed (red) or pollen (blue) parent pool. In total, 9,675 rye gene models carry one or more radical mutations, whereas 18,109 gene models carry no nsSNVs.



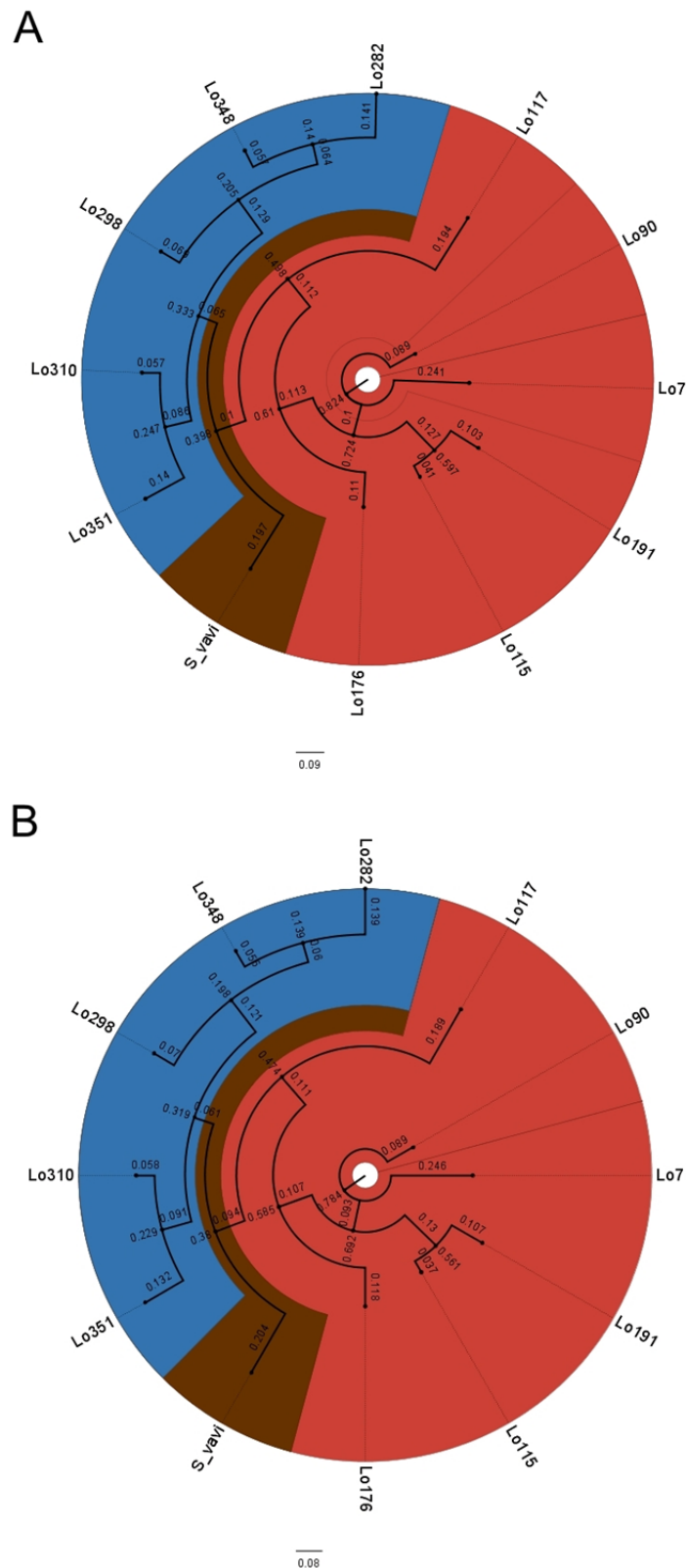
**Figure S3 Comparison of 2,934 gene model candidates showing presence/absence variation**

The diverging patterns of presence and absence observed in the rye gene-space between 11 resequenced genotypes (10 inbred lines and *S. vavilovii*) are represented by a heatmap for read coverage breadth ranging from yellow (full coverage of a gene model) to red (less than 5% of the gene length was covered).



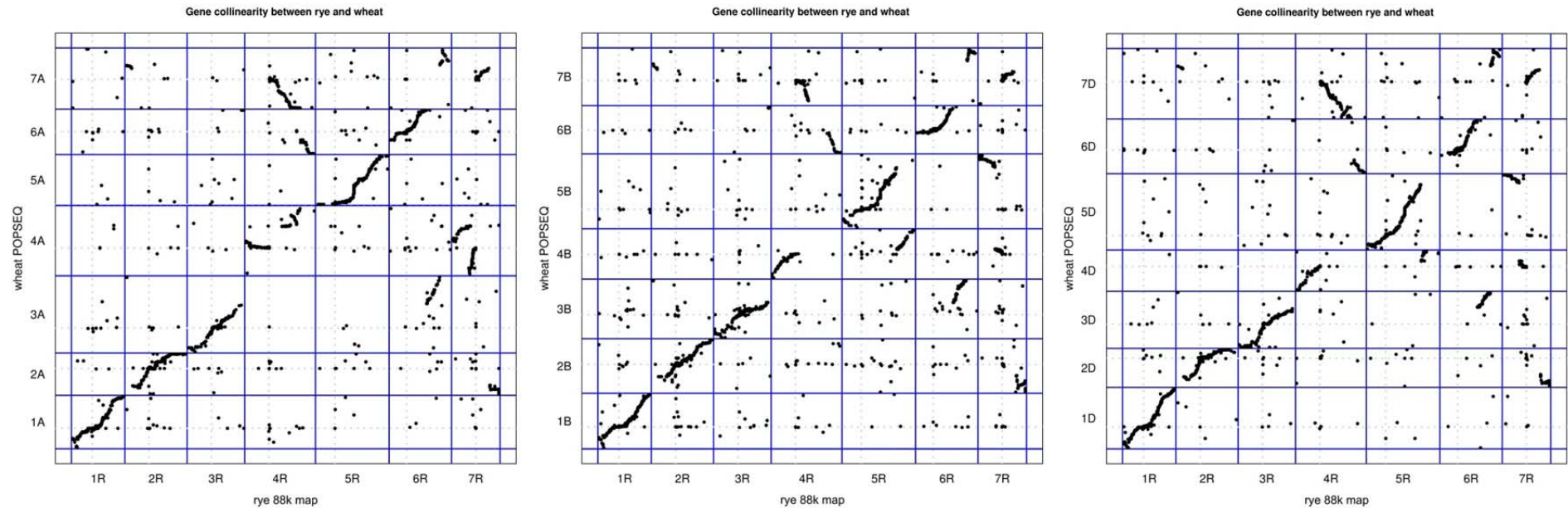
**Figure S4 Genotype proportion in the complete variant data set and the selected set of the Rye600k array**

The complete variant data set comprised 8,626,622 unique variant positions in the Lo7 assembly and the Rye600k variant data set 590,593 variant positions obtained from whole-genome sequencing. The bar plot shows which proportion (%) of the complete variant positions (blue) or the Rye600k variant positions (red) were detected in each genotype, i.e. where the genotype showed the alternative allele compared to the Lo7 reference allele. Absolute numbers of variants are given in the table below. The figure shows that the design of the Rye600k used a significant proportion of *S. vavilovii* SNVs to cover polymorphisms representative for rye genetic resources.



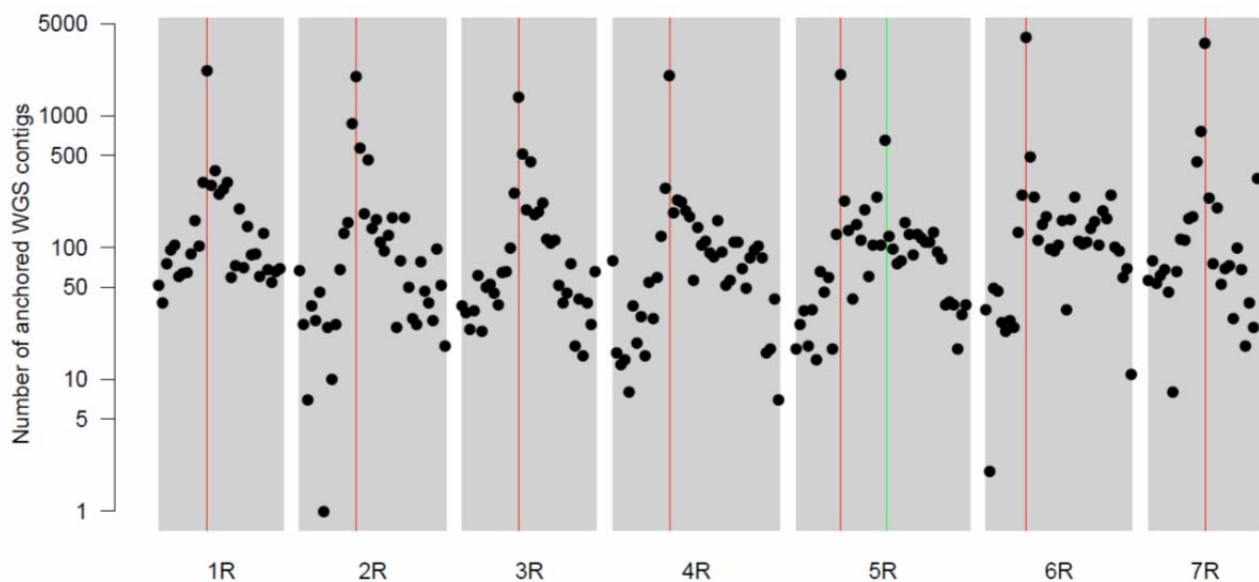
**Figure S5 Phylogenetic trees constructed (A) from complete set of chromosome-assigned SNVs and (B) from the Rye600k array**

The tree was constructed from concatenated WGS contigs of Lo7. The different rye breeding lines cluster in seed parent pool (red) and pollen parent pool (blue) and in between the close relative *S. vavilovii* (brown). Values at the branches were scaled in terms of expected numbers of substitutions. The scale was normalized to 1 defined as the average rate of changes using all changes observed in the dataset. For details see DNAML description (<http://evolution.genetics.washington.edu/phylip/doc/dnaml.html>).



**Figure S6 Collinearity between the genetic maps of rye and wheat**

The order of gene-bearing sequence contigs in the high-density genetic map of rye (x-axis) was compared to the order of their orthologous contigs in the assemblies of the three wheat subgenomes A, B and D (Mayer et al. 2014) (y-axis). Chromosomes are separated by blue lines. Positions of genetic centromeres are marked with dotted grey lines.



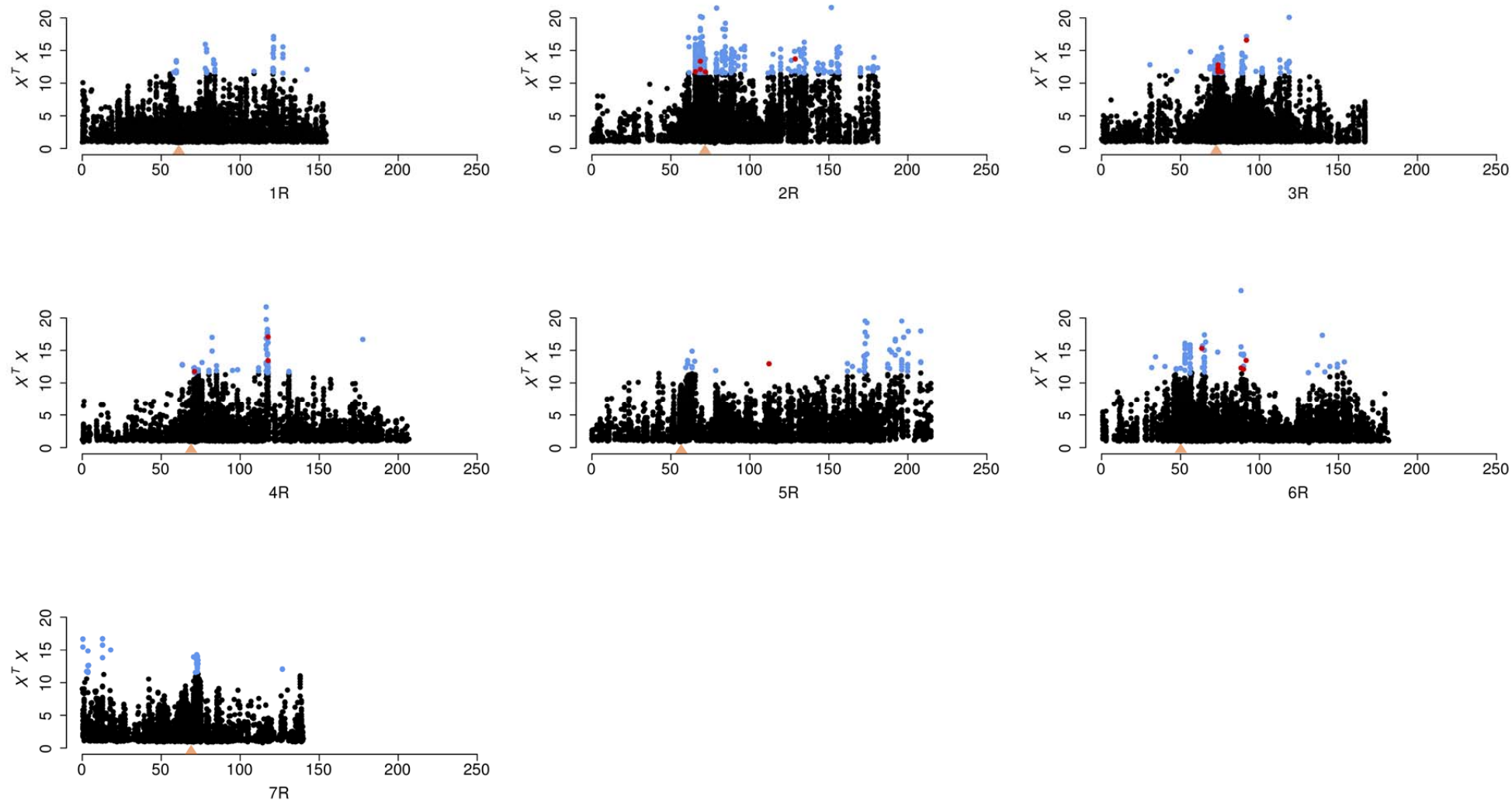
Chromosome	1R	2R	3R	4R	5R	6R	7R
Bin (5 cM)	60	70	70	70	55	50	70
# contigs (5 cM)	2,198	2,001	1,392	2,010	2,076	3,962	3,557
Bin (1 cM)	61	72	75	71	56	50	70
# contigs (1 cM)	1,801	1,241	1,182	1,445	891	3,972	3,715

**Figure S7 Position of genetic centromeres in the rye genetic map**

Top: The number of anchored WGS contigs (y-axis) were tabulated in 5 cM bins and plotted along the genetic map (x-axis). The position of 5 cM bins with the highest number of anchored contigs was denoted as the position of the genetic centromere and marked with a red line. A second region of suppressed recombination on chromosome arm 5RL, putatively caused by neocentric activity, is marked with a green line.

Bottom: The positions of the genetic centromeres were determined for bin sizes of 5 cM and 1 cM as the bin with the largest number of anchored sequence contigs (see table).

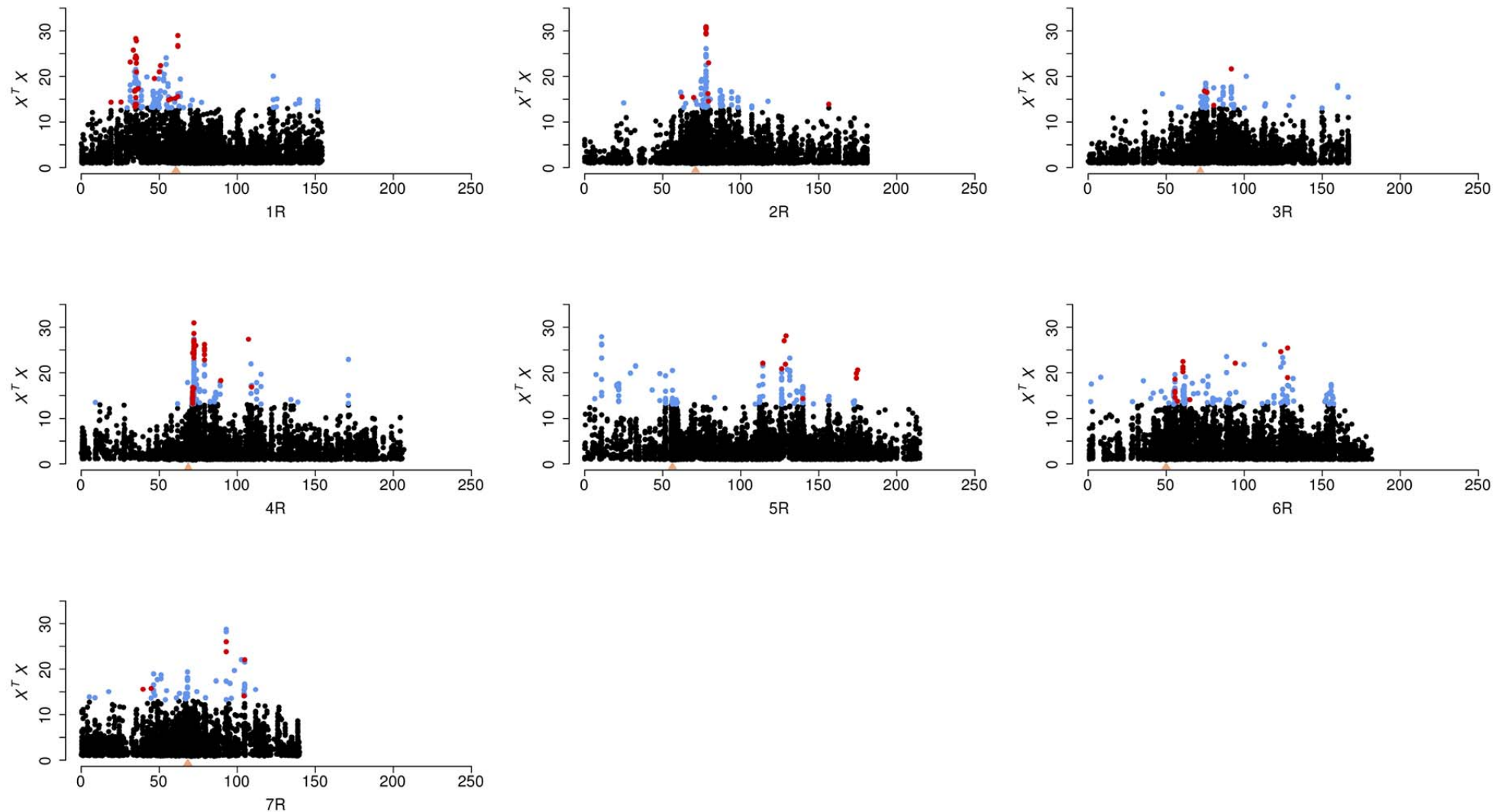




**Figure S8 Genome-wide map of selection signals between the seed parent pool and genetic resources**

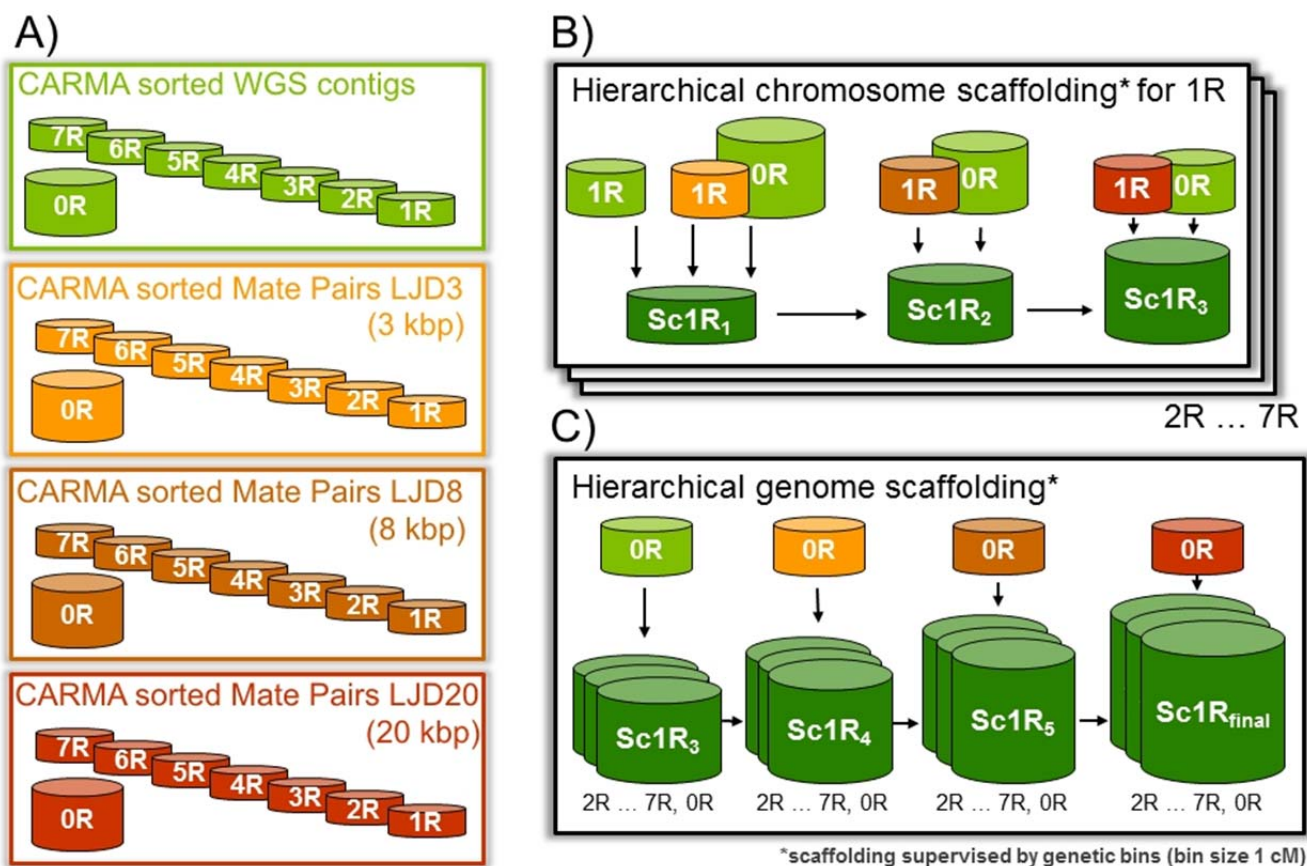
The plots for the seven rye chromosomes are based on 78,731 genetically mapped SNVs from the Rye600k array. The blue and red SNVs are the top 1%  $X^T X$  values. The red SNVs are shared Lositan ( $F_{ST}$ ) outliers and top 1%  $X^T X$  values. Centromere positions are indicated by the triangle on the x-axis.





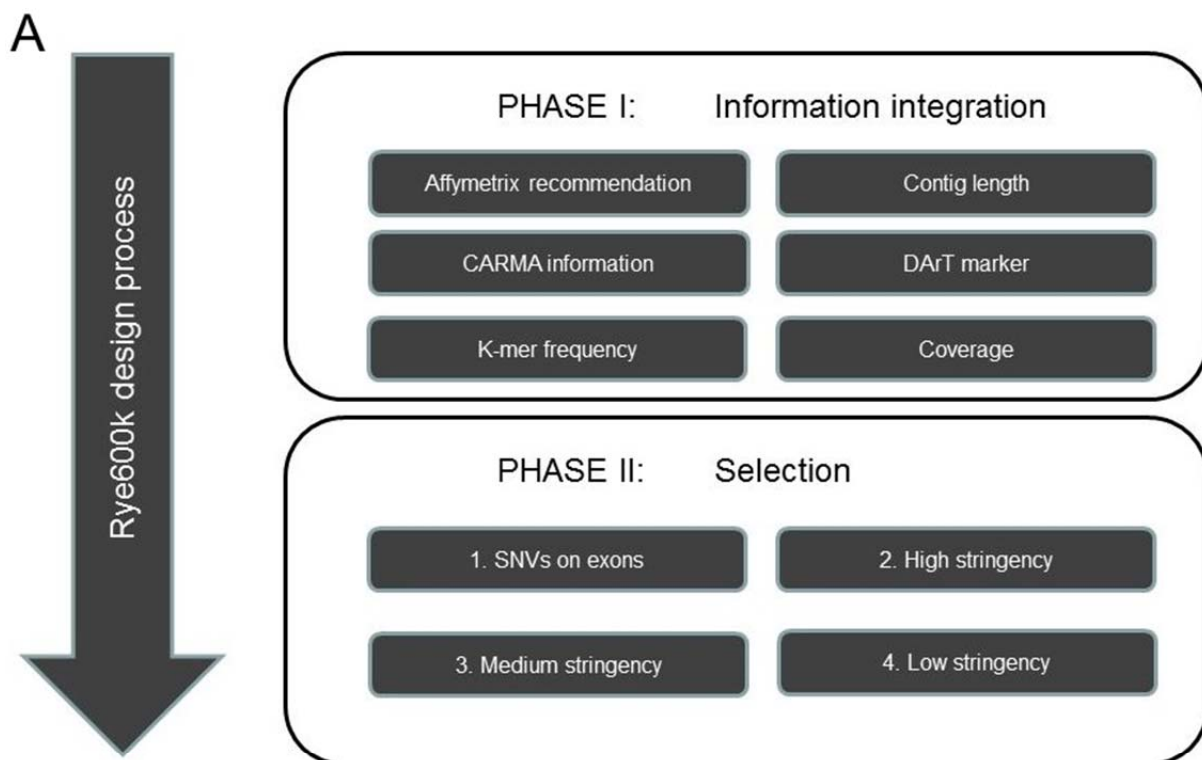
**Figure S9 Genome-wide map of selection signals between the pollen parent pool and genetic resources**

The plots for the seven rye chromosomes are based on 78,731 genetically mapped SNVs from the Rye600k array. The blue and red SNVs are the top 1%  $X^T X$  values. The red SNVs are shared Lositan ( $F_{ST}$ ) outliers and top 1%  $X^T X$  values. Centromere positions are indicated by the triangle on the x-axis.



### Figure S10 Hierarchical scaffolding scheme

The scaffolding process was divided into the three sub-processes: A) Chromosome sorting of WGS contigs and mate-pair (MP) libraries using CARMA. B) Chromosome-specific scaffolding using a hierarchical approach that started with short distance (LJD3, 3 kbp) MP libraries and subsequently inserted the remaining libraries with increasing distances (LJD8, 8 kb and LJD20, 20 kbp). The sub-process was performed for each chromosome individually. C) Genome scaffolding applying the hierarchical approach using the complete set (assigned and un-assigned sequences) of scaffolded and un-scaffolded WGS contigs. To avoid chimeric sequences, the scaffolding process was supervised using the information of a genetic map. With this control the scaffolding of WGS contigs that had a genetic distance  $>1\text{cM}$  was not permitted.



**B**

Parameter	High stringency threshold	Medium stringency threshold	Low stringency threshold
Min. number of genotypes showing alternative allele	4	2	1
Max. number of genotypes showing alternative allele	10	10	10
Min. contig length (bp)	1,500	1,000	500
Min. number of genotypes with read coverage at SNV position	4	3	2
Avg. K-mer frequency of probe sequence below threshold	10	10	10
Affymetrix recommendation	recommended	recommended or neutral	recommended or neutral
Max. number of marker per contig (N = contig length / threshold)	1,000	1,000	1,000
Distance to contig end (bp)	100	100	100

**Figure S11 Process and parameter settings of the Rye600k array design**

A) The process was divided in two phases. Phase I aggregated a wide spectrum of pre-computed information and the phase II combined these features of a SNV to select markers with the best priority setting. In total 8.6 million SNVs were analysed that all fulfilled the prior requirements of a qualitative variant call. At the end of the evaluation 600,843 markers were represented on the Rye600k array (including 10,250 markers from a previous custom Illumina Infinium SNP array).

B) Parameter settings for selection of SNVs for the Rye600k array using three stringency thresholds.