Plant Communications, Volume *3*

Supplemental information

A recent burst of gene duplications in Triticeae

Xiaoliang Wang, Xueqing Yan, Yiheng Hu, Liuyu Qin, Daowen Wang, Jizeng Jia, and Yuannian Jiao

Supplemental Information

A recent burst of gene duplications in Triticeae

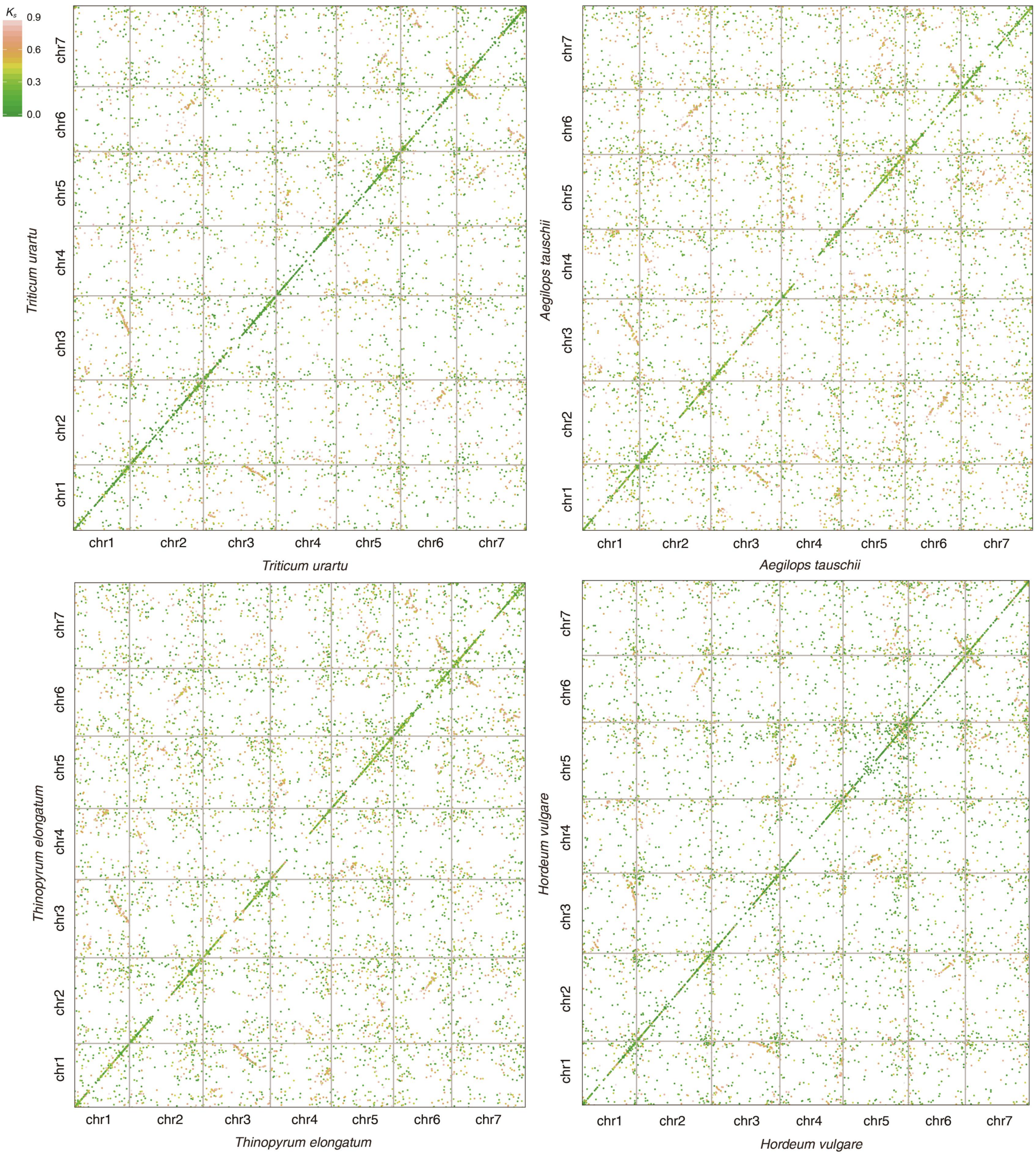
Xiaoliang Wang[1,2#], Xueqing Yan[1,2#], Yiheng Hu[1,2#], Liuyu Qin[1,2], Daowen Wang[3*], Jizeng Jia[3,4*], Yuannian Jiao[1,2*]

[1]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China.

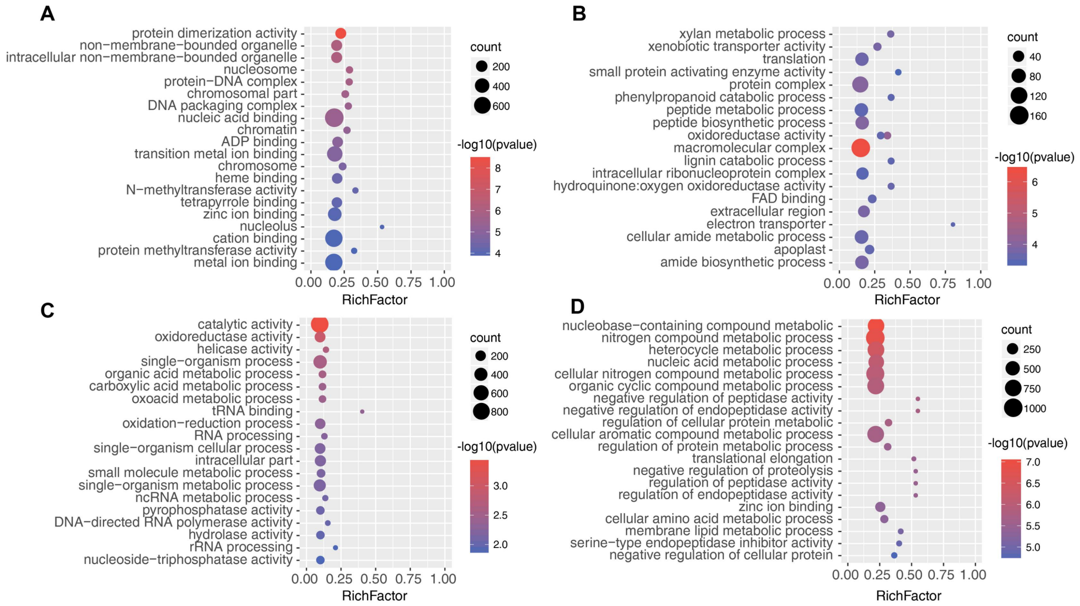
[2]University of Chinese Academy of Sciences, Beijing 100049, China.

[3]College of Agronomy, Collaborative Innovation Center of Henan Grain Crops, Henan Agricultural University, Zhengzhou, Henan 450046, China.

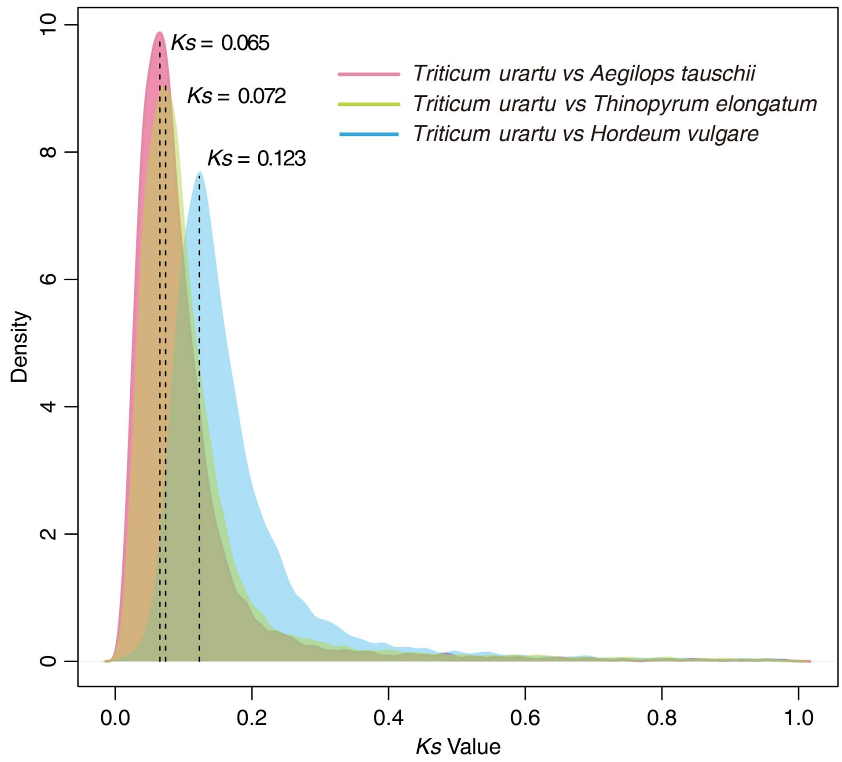
[4]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China.



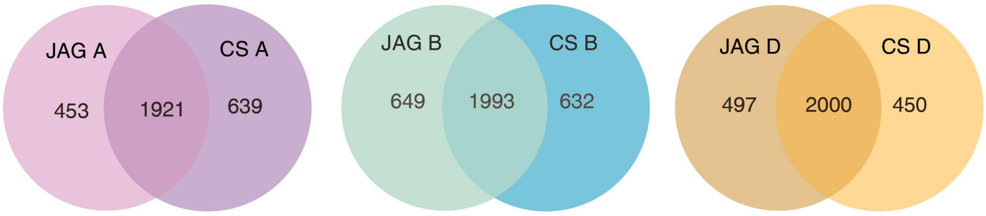
Supplementary Figure 1. Dot plots of genomic or subgenomic self-comparisons. Intra-genomic dot plots of gene pairs retrieved from all against all best reciprocal hits in *T. urartu, Ae. tauschii*, *Th. elongatum* and *H. vulgare* genome respectively. K_s value of each gene pair was shown by plotting different colors.



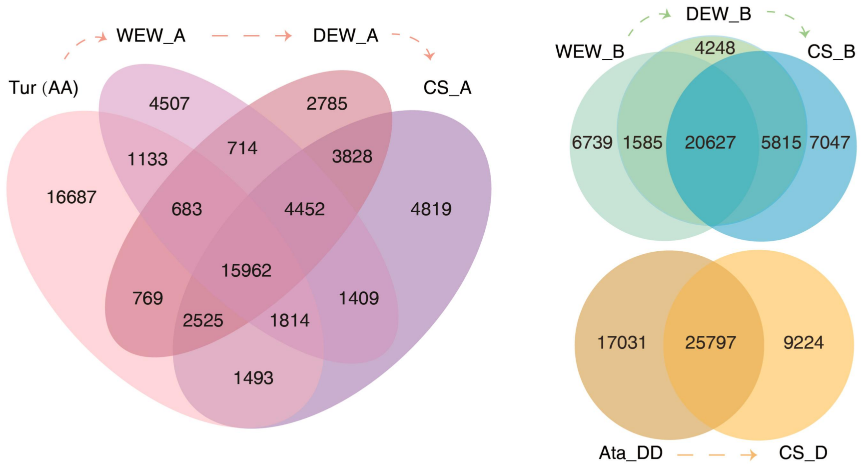
Supplementary Figure 2. Significantly enriched GO terms for the recently duplicated genes in (A) *T. urartu,* **(B)** *Ae. tauschii*, **(C)** *Th. elongatum* and **(D)** *H. vulgare* genome. The results are sorted according to significance, of which GO term of protein dimerization activity, xylan metabolic process, catalytic activity, and nucleobase-containing compound metabolic are the most significant respectively.



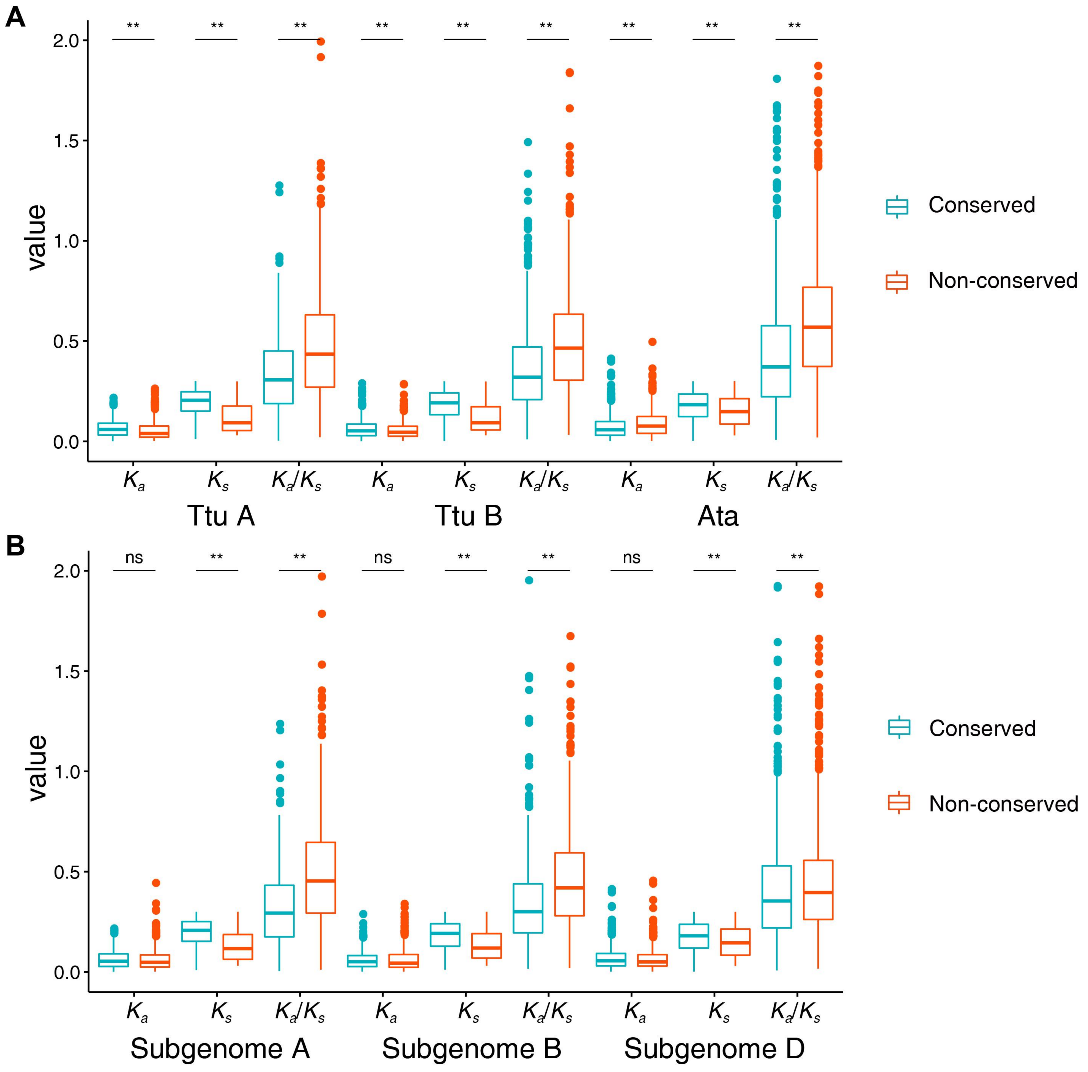
Supplementary Figure 3. Density plot of K_s values of the best reciprocal hits comparing *T. urartu* to *Ae. tauschii*, *Th. elongatum* and *H. vulgare* respectively. The three K_s values correspond to the differentiation of the Triticeae with the K_s around 0.123, the differentiation of *Th. elongatum* and *Triticum* with the K_s around 0.072, and the differentiation of *Triticum* with the K_s around 0.065 respectively.



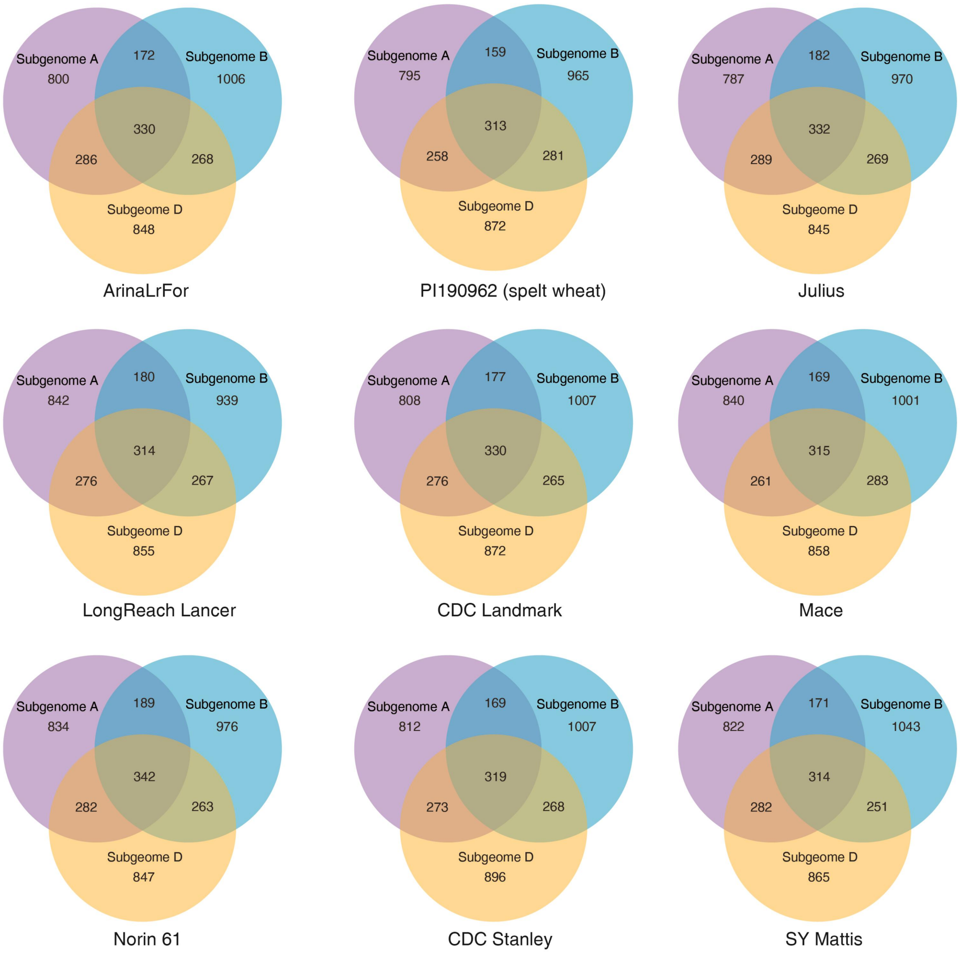
Supplementary Figure 4. Venn diagrams show commonly-retained and specific-retained recent duplicates in wheat cultivars of JAG and CS.



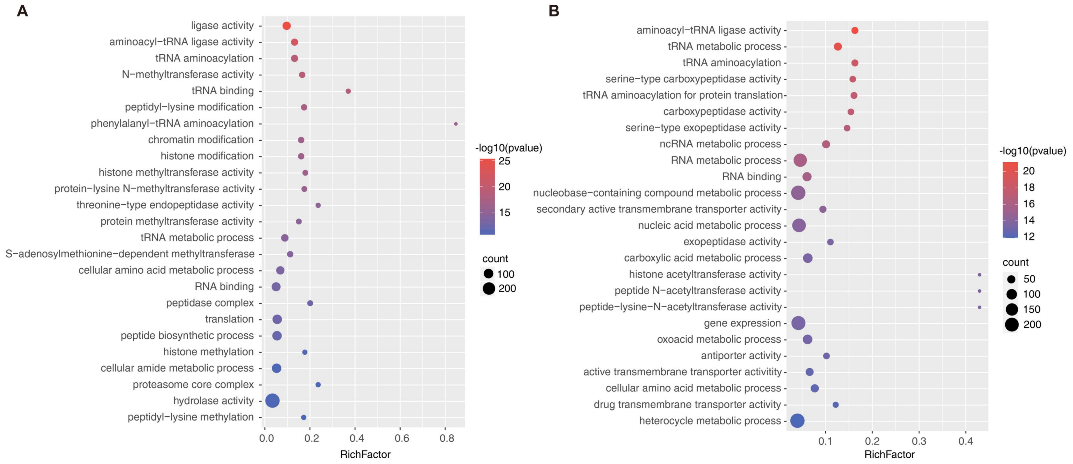
Supplementary Figure 5. Venn diagram shows the numbers of orthologous in CS subgenomes and their progenitor genomes.



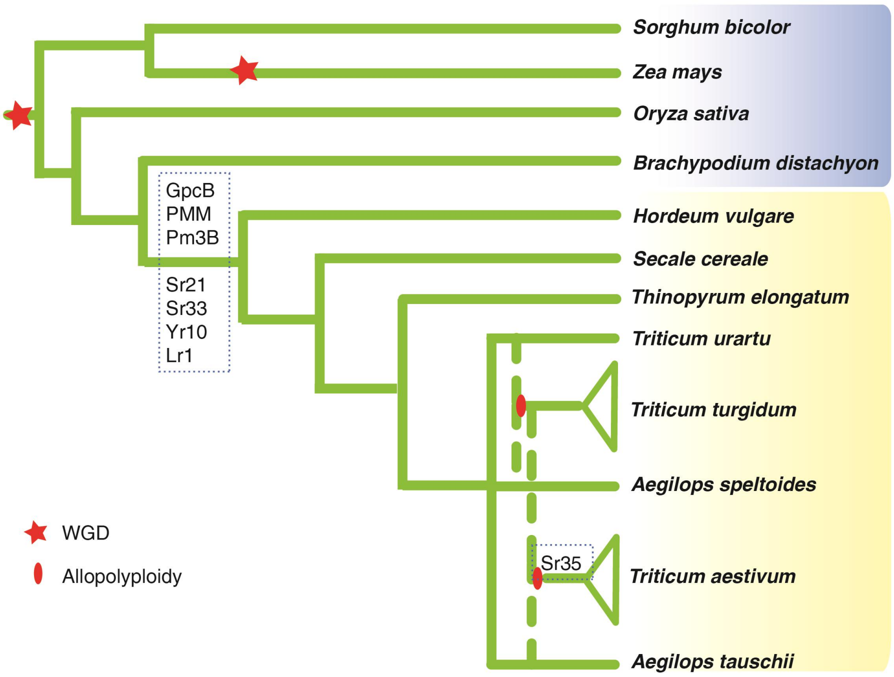
Supplementary Figure 6. Sequence divergence and selection analyses of the orthologous gene pairs between CS and its progenitor species. (A) The K_a, K_s, and K_a/K_s analyses of recent duplicates in progenitor genomes. Conserved gene pairs refer to gene pairs in the progenitor genomes requiring both genes have corresponding genes in CS, while non-conserved ones mean the gene pairs that have no corresponding genes in CS. **(B)** The K_a, K_s, and K_a/K_s analyses of recent duplicates in three subgenomes of CS. Conserved gene pairs refer to gene pairs that have corresponding genes with the progenitor genomes of CS, while non-conserved ones mean those that are not corresponding with the progenitor genome of CS. ns (not significant) $P > 0.05$, *$P < 0.05$, ** $P < 0.01$ in Wilcoxon test.



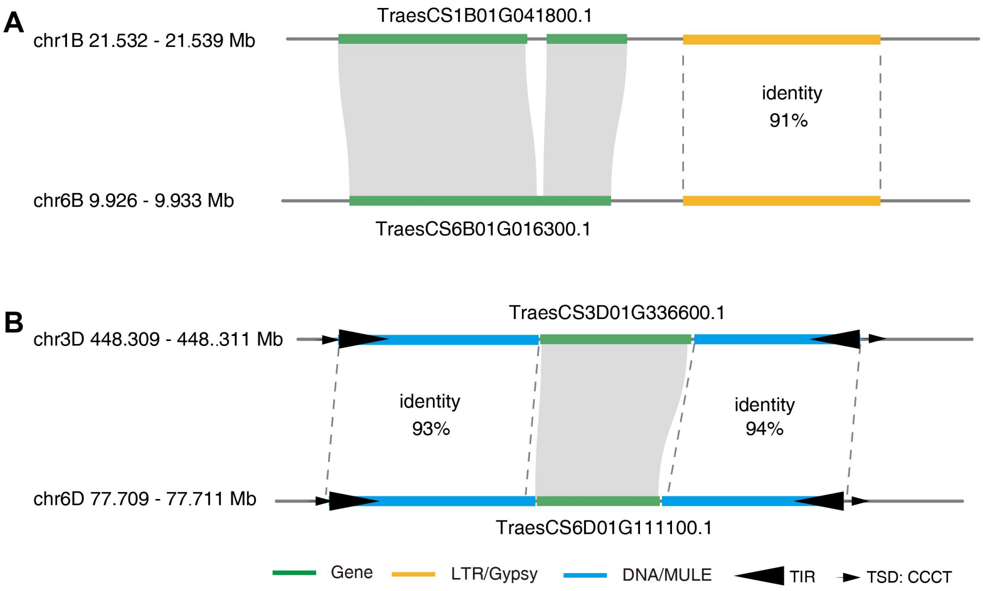
Supplementary Figure 7. Venn diagram shows the commonly retained recent gene duplicates for the three subgenomes of the nine wheat genomes.



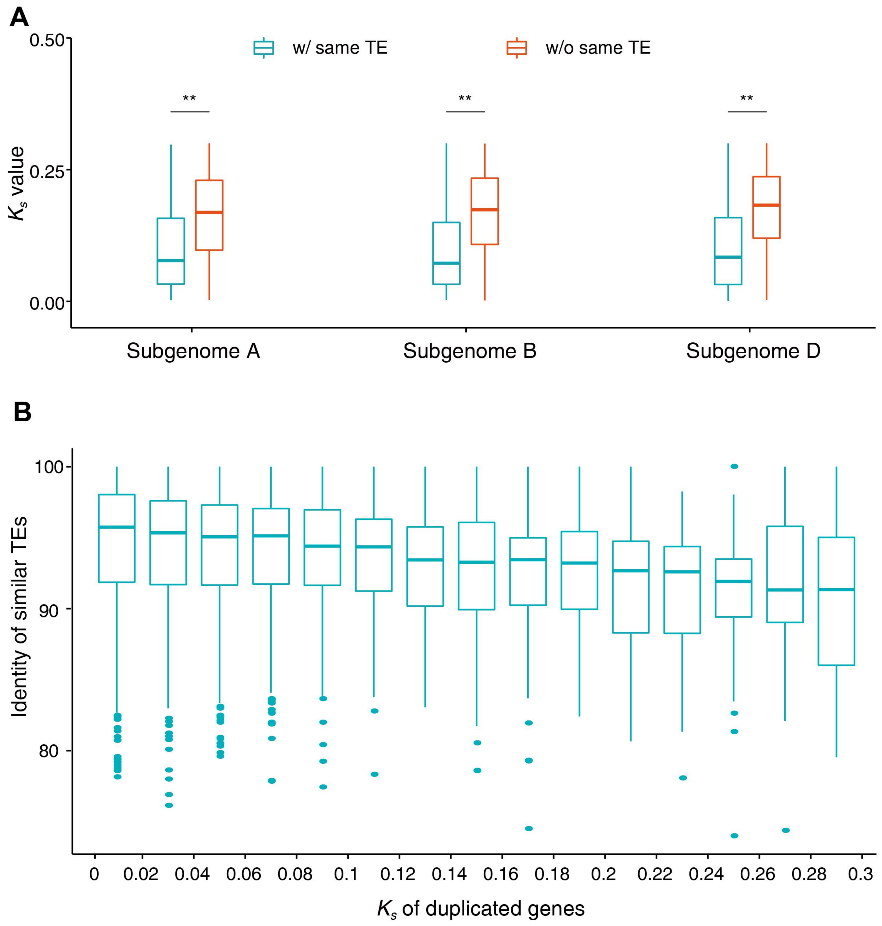
Supplementary Figure 8. Significantly enriched GO terms for well retained genes in the three subgenomes of (A) CS and (B) JAG. The results are sorted according to the significance, of which GO term of ligase activity is the most significant.



Supplementary Figure 9. Duplication timing of the eight previously identified agronomically important genes in the RBGD.



Supplementary Figure 10. Two examples of gene duplication potentially derived from TE activity in CS genome. Identity of gene pairs is greater than 90%. **(A)** The genes of *TraesCS1B01G041800.1* and *TraesCS6B01G016300.1* locate beside TEs of the same subtype and with 91% sequence identity; the duplicated copy (*TraesCS6B01G016300.1*) exhibits intron-less. **(B)** *TraesCS3D01G336600.1* and *TraesCS6D01G111100.1* locate within TEs of the same subtype (DNA/MULE) and with 94% sequence identity. TIR: terminal inverted repeat, TSD: target site duplication.



Supplementary Figure 11. Evolutionary rate of recent duplicate gene pairs in CS. **(A)** K_s of duplicates that are either flanked or not flanked by a given TE type. ** $P < 0.01$ in Wilcoxon test. **(B)** Sequence identity of similar TEs during the evolution process. The abscissa are K_s values in each 0.02 window of recently duplicated genes with the same subtype of TE.