## Supporting Figures

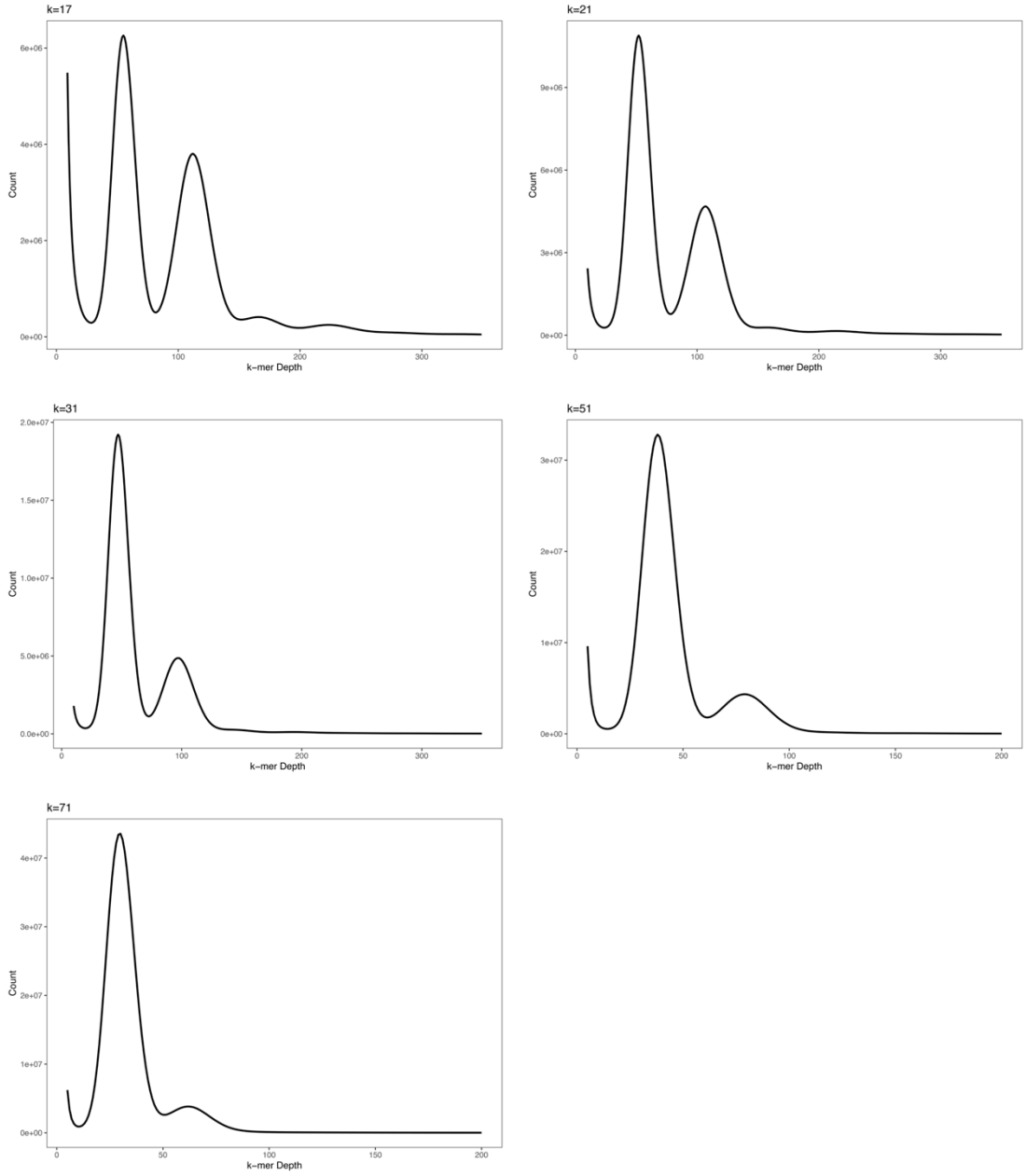



**Figure S1** K-mer based genome size estimates. K = 17, 21, 31, 51 and 71 count histograms of Illumina paired-end sequence data for tetraploid *Trapa natans*. These plots were used to estimate genome size (total k-mer number)/(peak depth). K = 17: a total of 50,029,871,544 17-mers were detected with a peak depth of 54 and an estimated genome size of 926.48 Mb; K = 21: a total of 51,476,584,112 21-mers were detected with a peak depth of 52 and an estimated genome size of 989.93 Mb; K = 31: a total of 50,607,610,875 31-mers were detected with a peak depth of 47 and an estimated genome size of 1076.76 Mb; K = 51: a total of 43,270,633,852 51-mers were detected with a peak depth of 38 and an estimated genome size of 1138.70 Mb; K = 71: a total of 34,807,248,461 71-mers were detected with a peak depth of 30 and an estimated genome size of 1160.24 Mb.

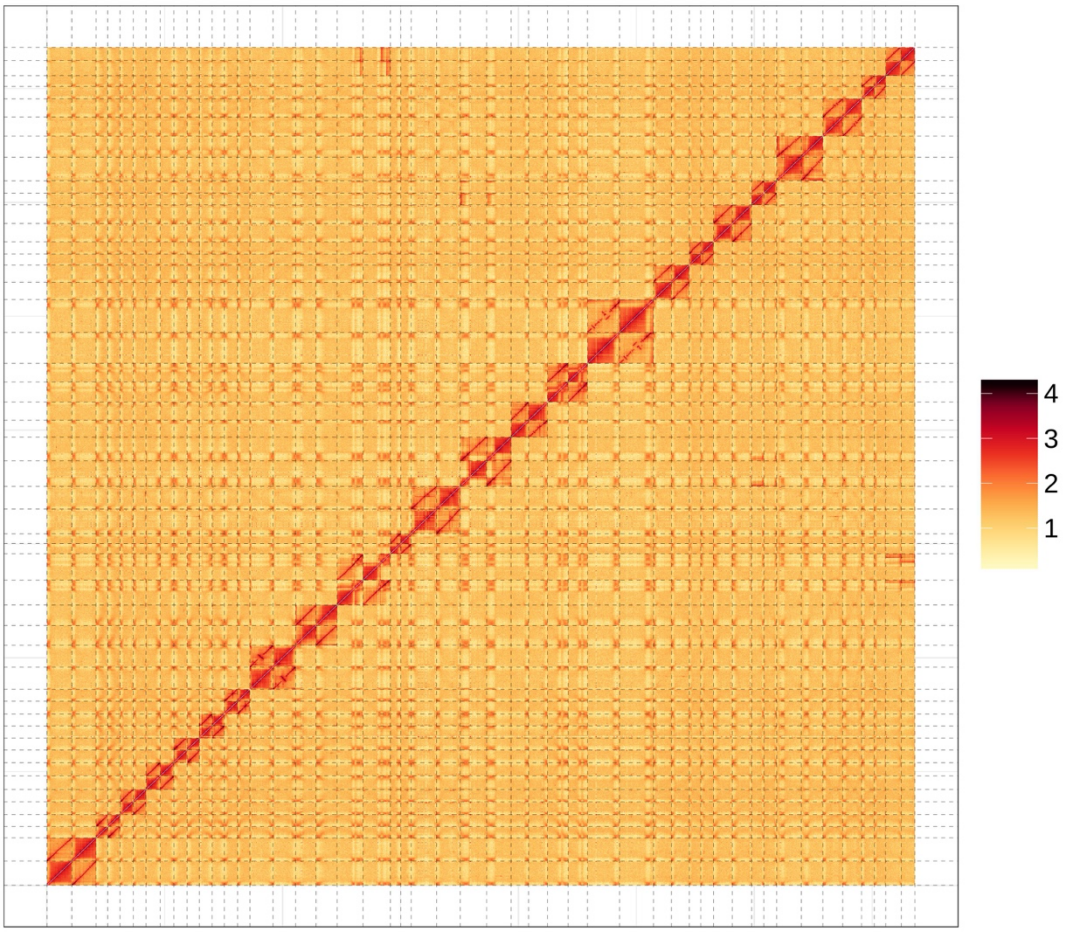



**Figure S2** Genome-wide analysis of chromatin interactions in the tetraploid *Trapa natans* genome based on Hi-C data.

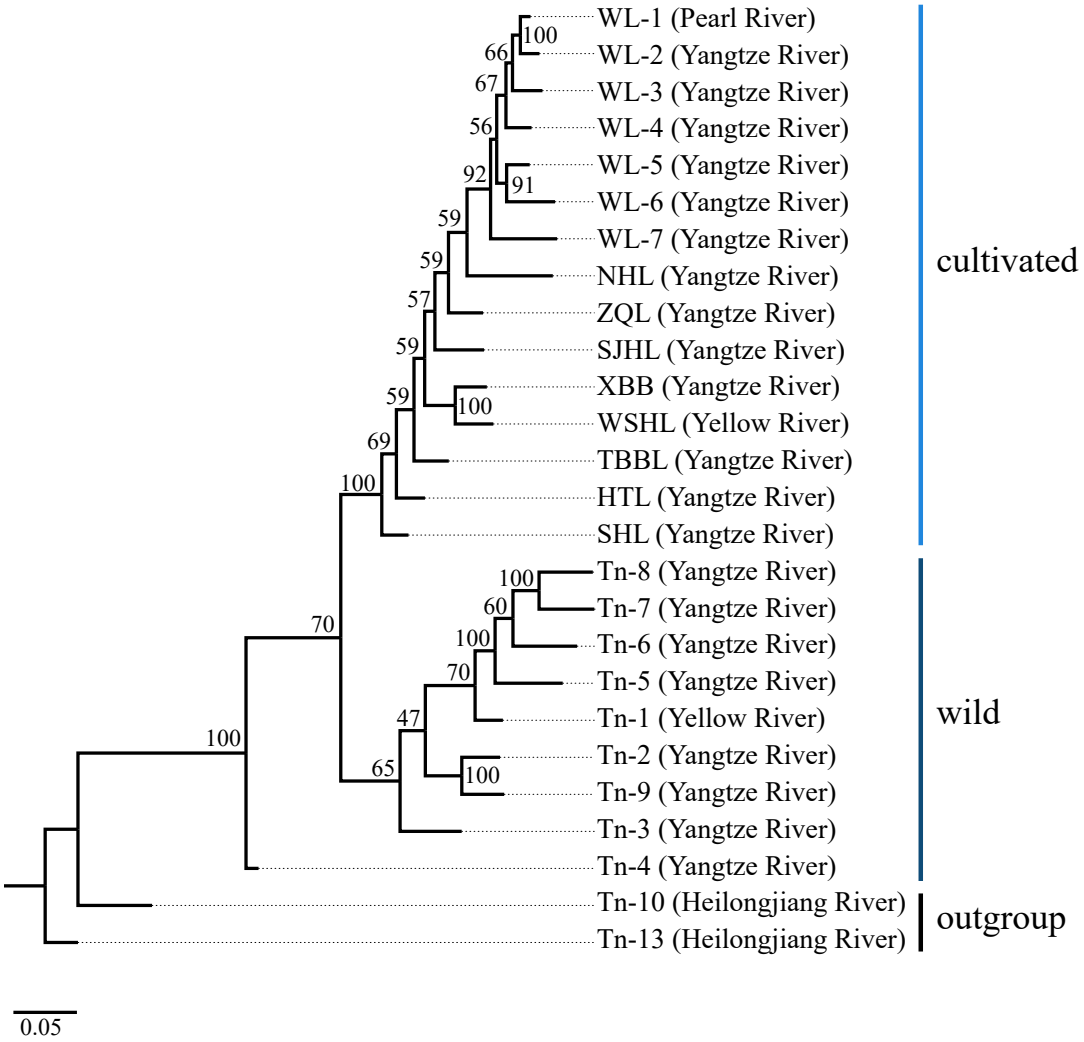



**Figure S3** Phylogenetic relationships of diploid *Trapa natans* (2x, AA) from the PYY region inferred from maximum likelihood (ML) analysis based on SNP data.

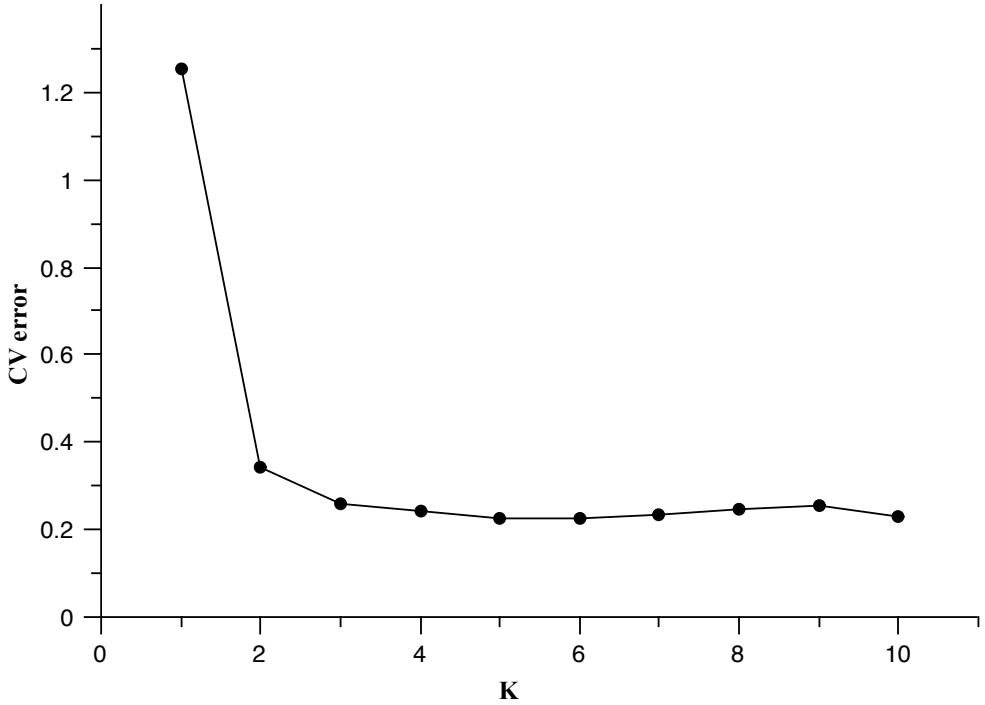



**Figure S4** Plot of ADMIXTURE cross-validation error across values of K.

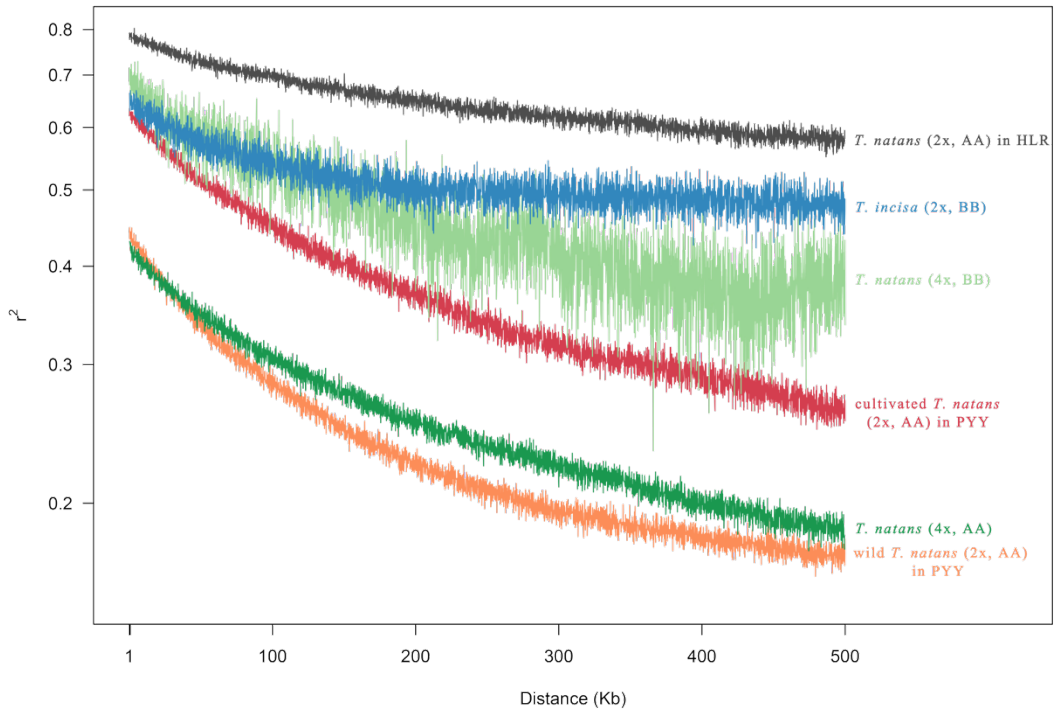



**Figure S5** LD decay determined by correlation of allele frequencies ($r^2$) against distance (kb) in six *Trapa* populations. Note that the A and B subgenomes of tetraploid *T. natans* individuals were treated separately (4x, AA vs. BB). PYY, Pearl-Yangtze-Yellow River region; HLR, Heilongjiang River region.

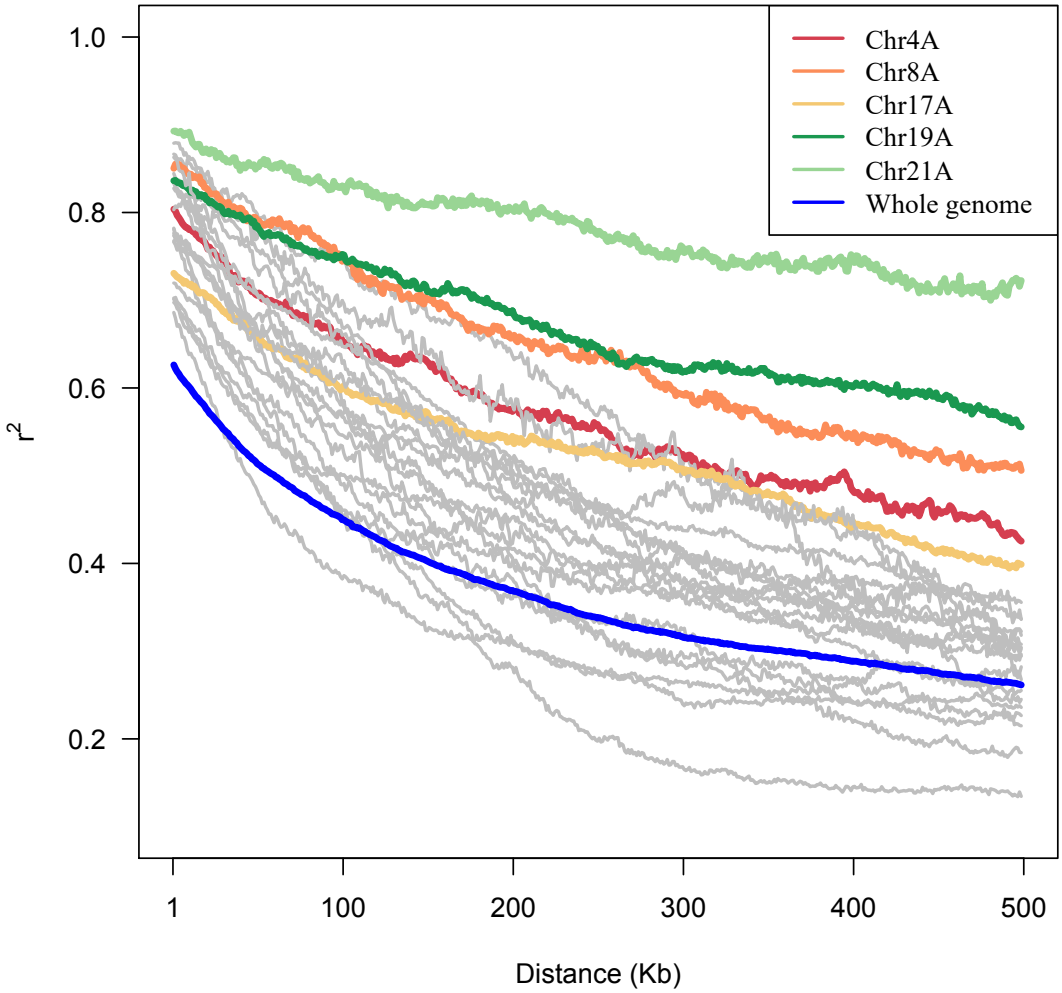



**Figure S6** LD decay distances in cultivars of *Trapa natans* for each of the 24 chromosomes.

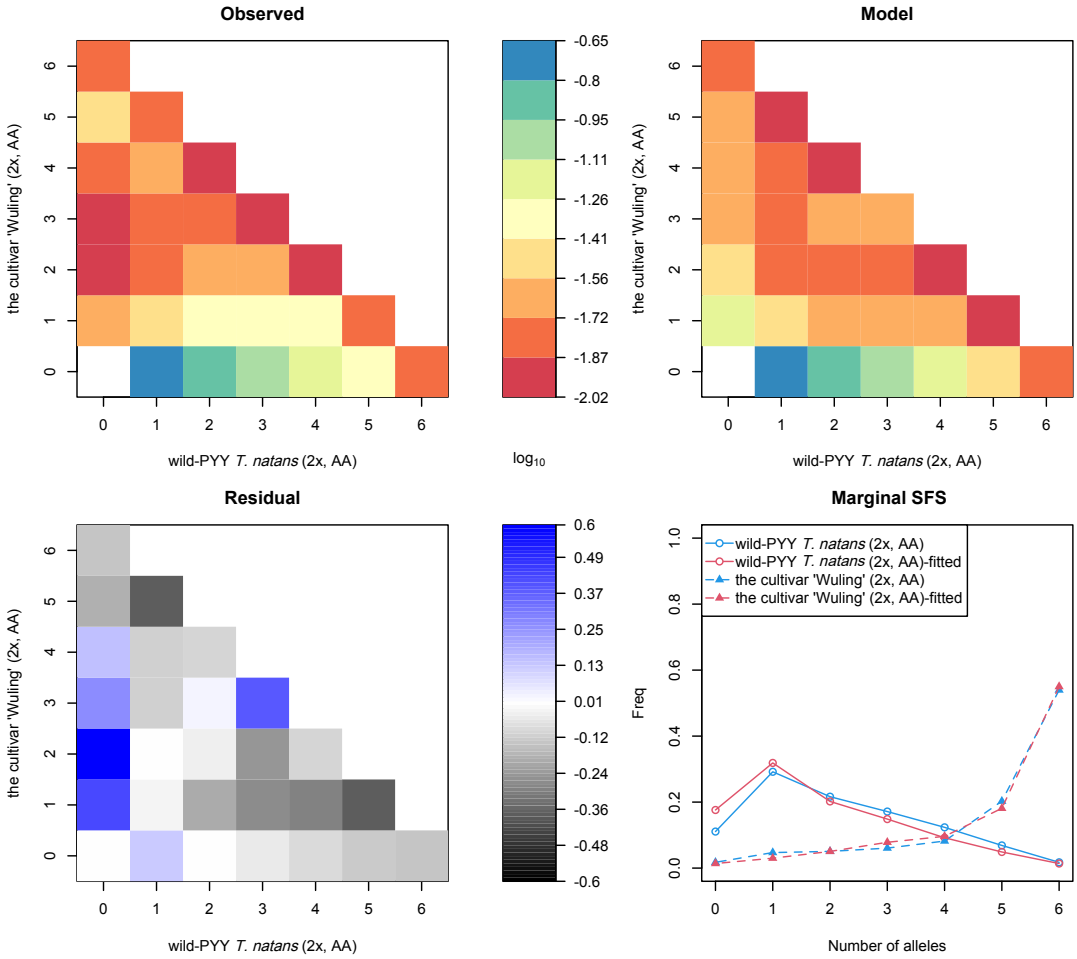



**Figure S7** The goodness-of-fit of the best scenario summarized by the observed 2D-SFS, modeled 2D-SFS, residuals between the model and the data, and marginal SFS.

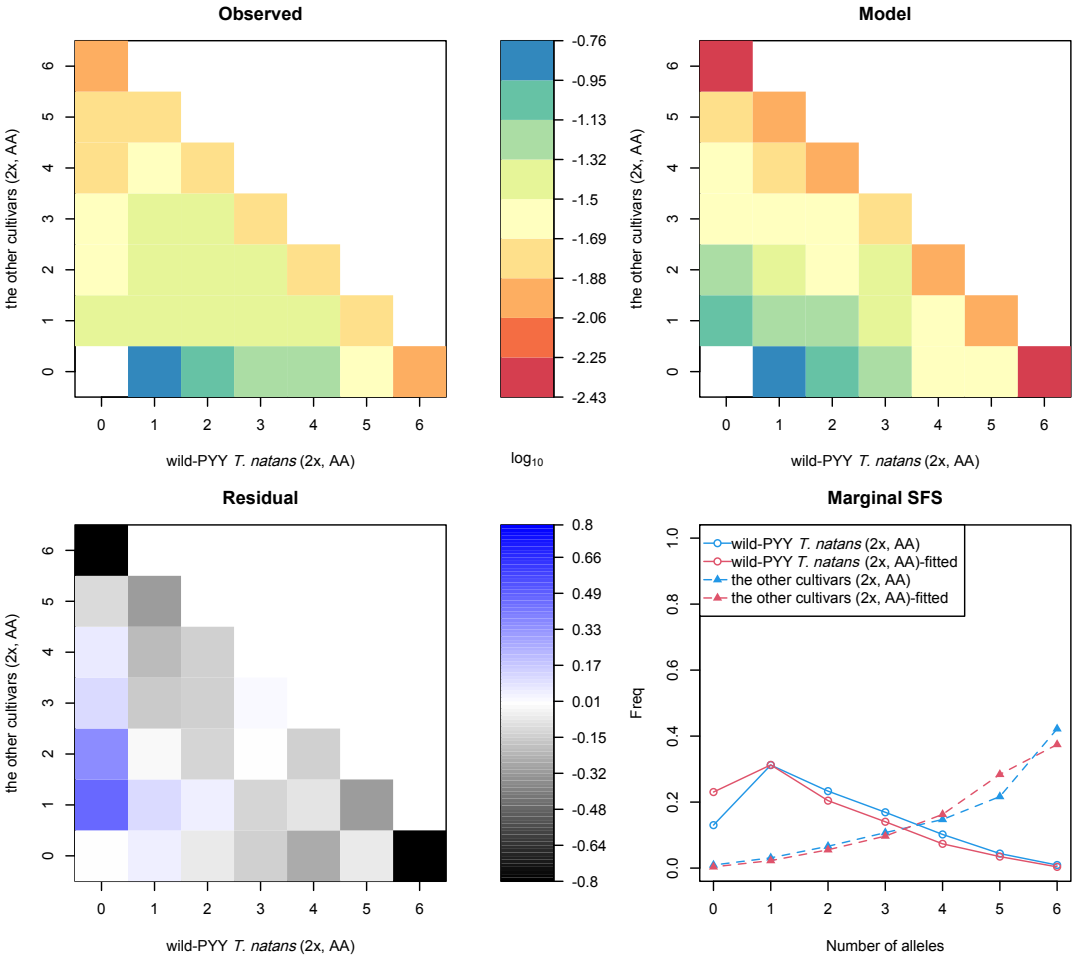



**Figure S7 (continued)**

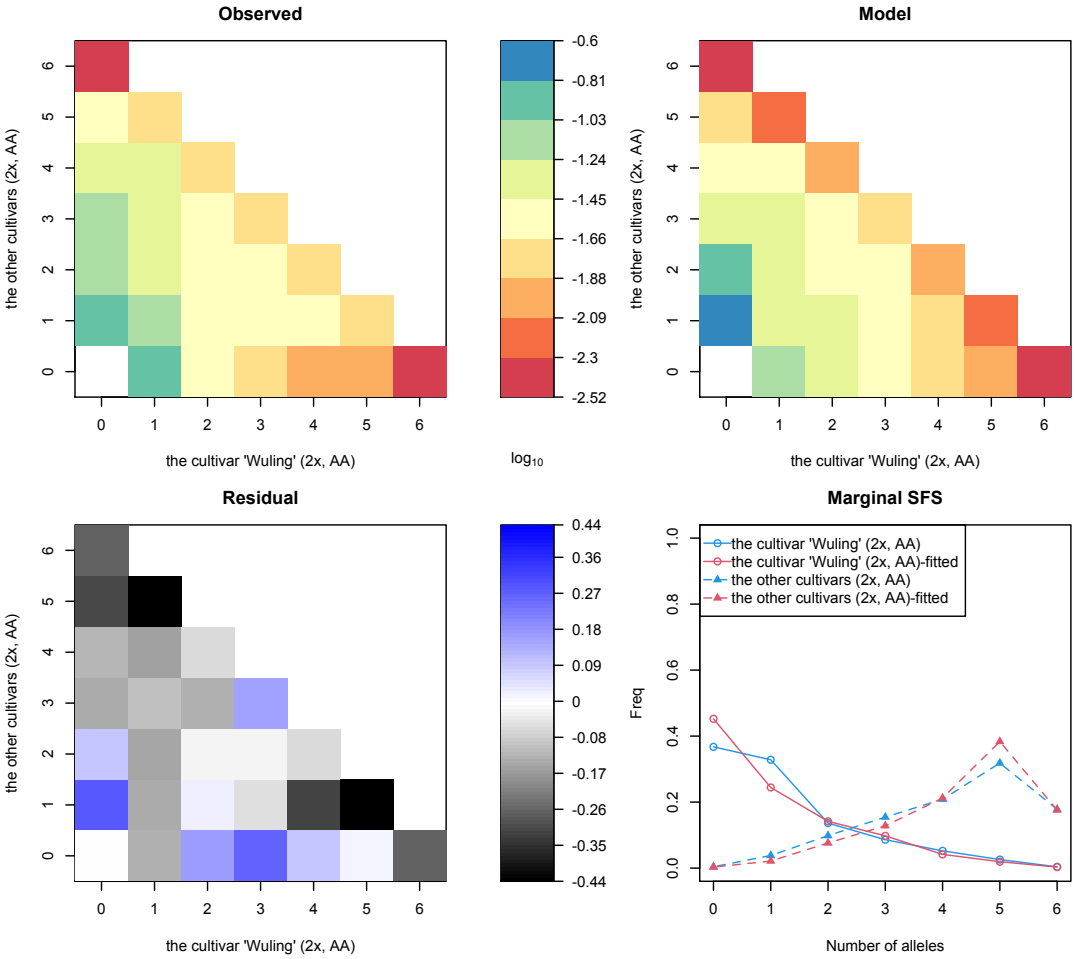



**Figure S7 (continued)**

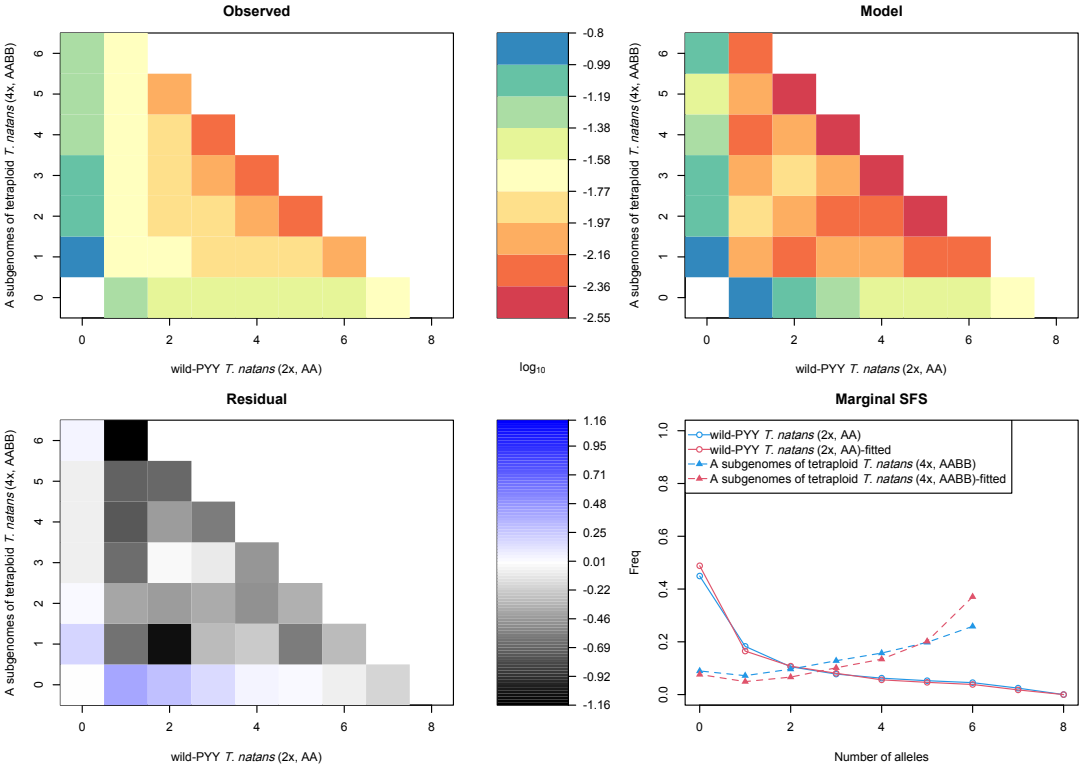



**Figure S7 (continued)**

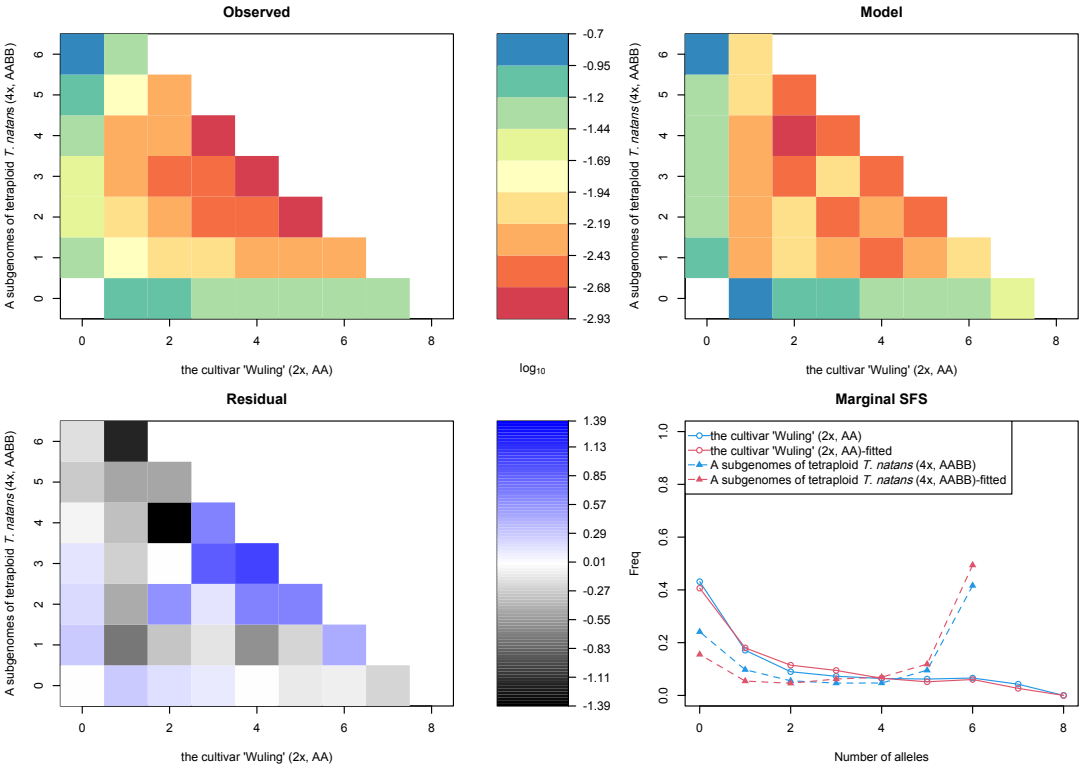



**Figure S7 (continued)**

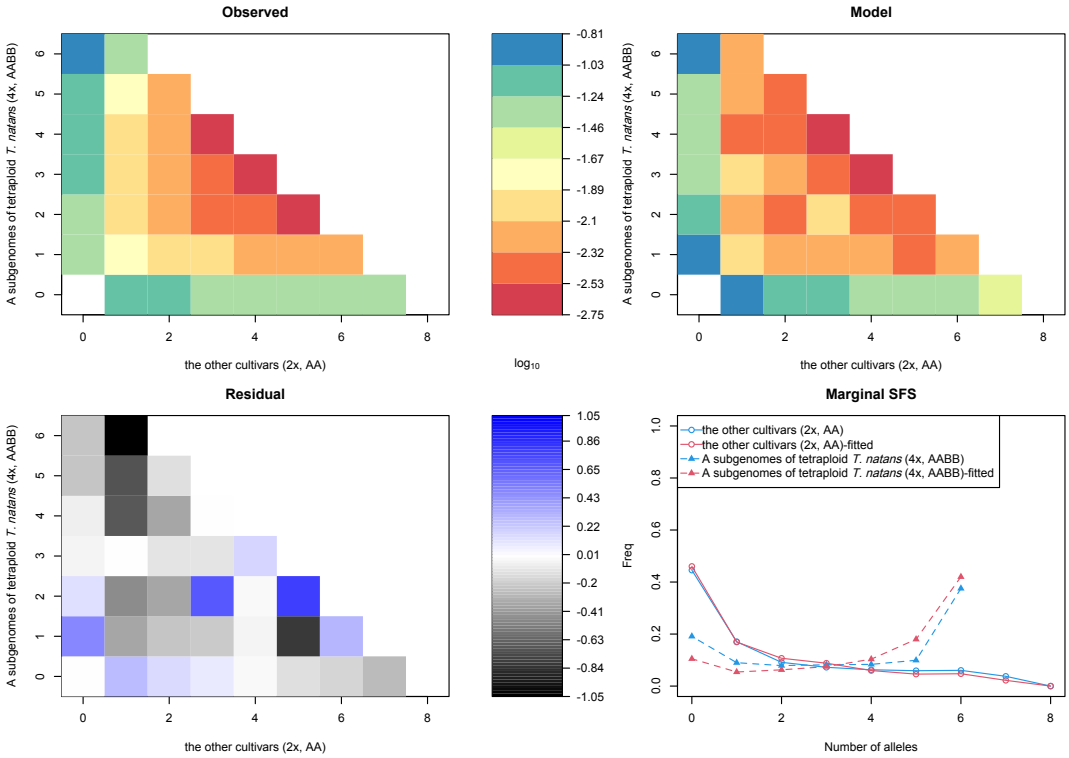



**Figure S7 (continued)**

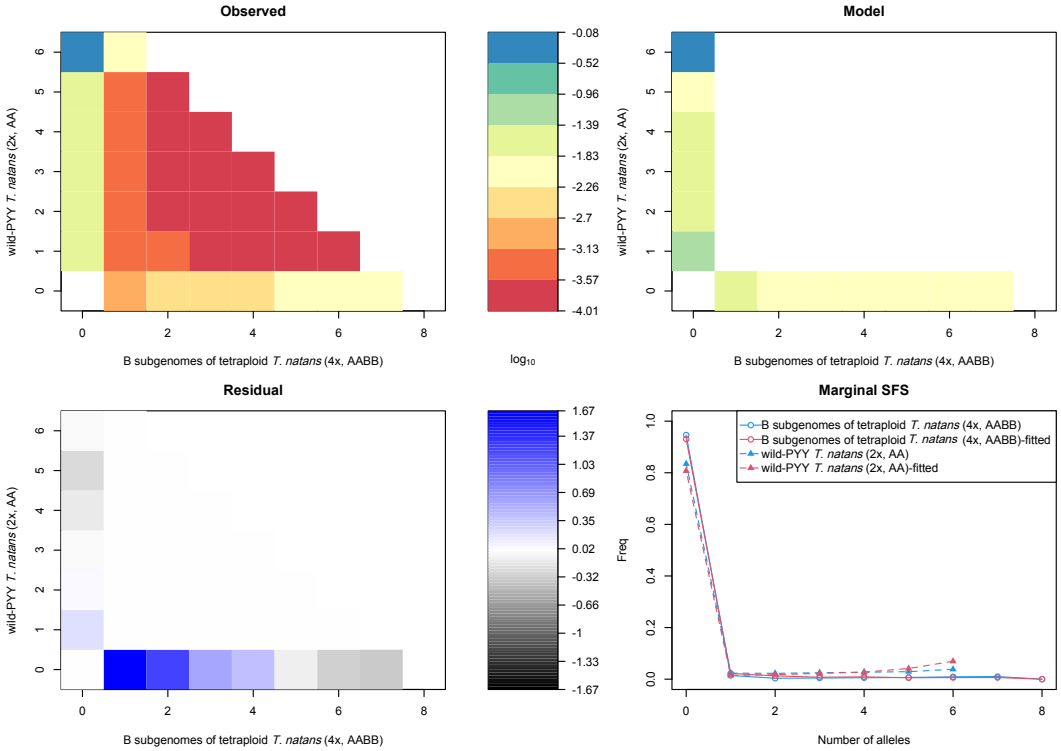



**Figure S7 (continued)**

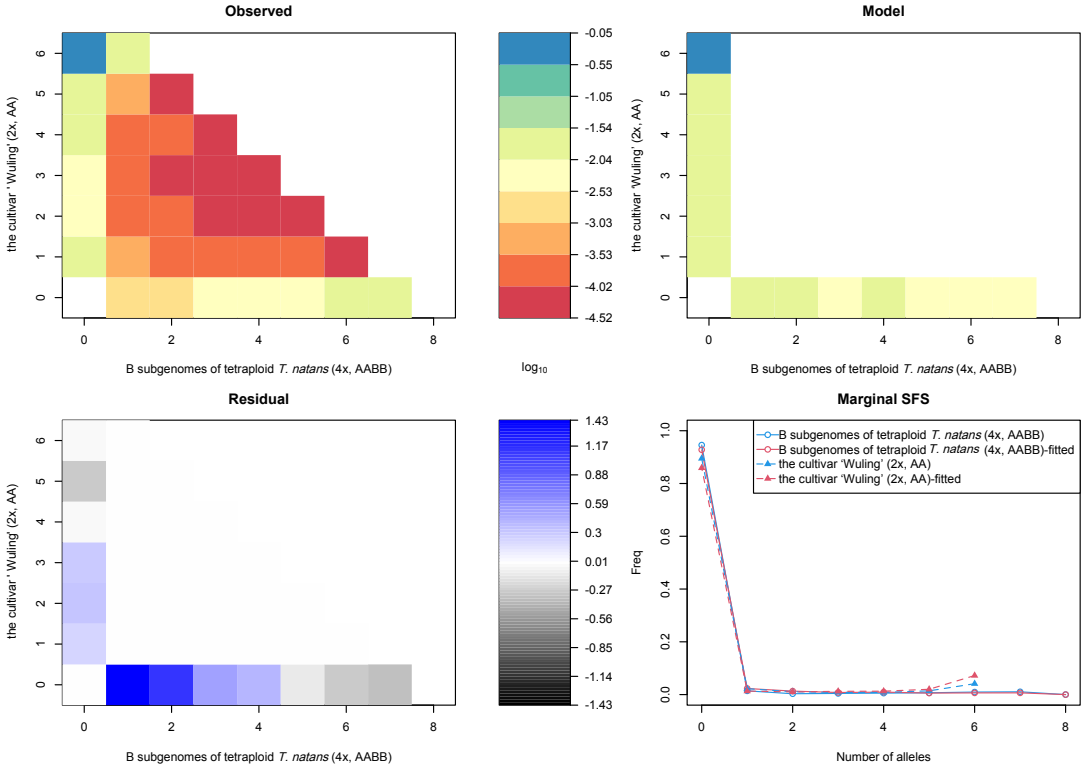



**Figure S7 (continued)**

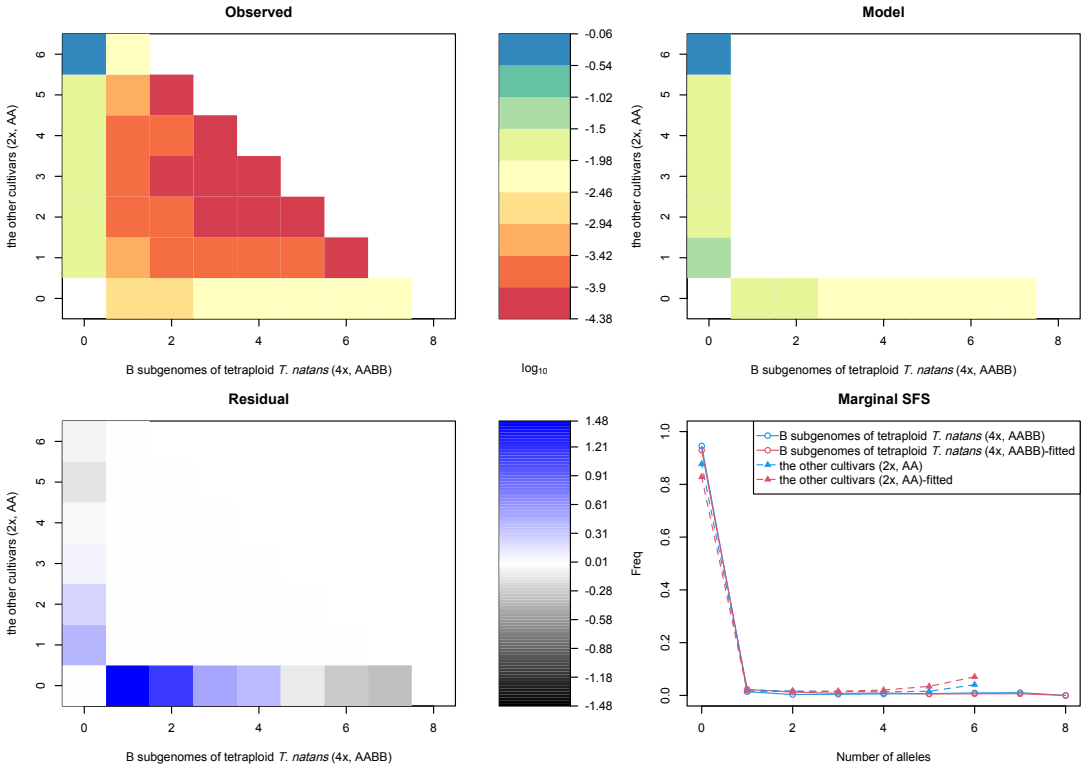



**Figure S7 (continued)**

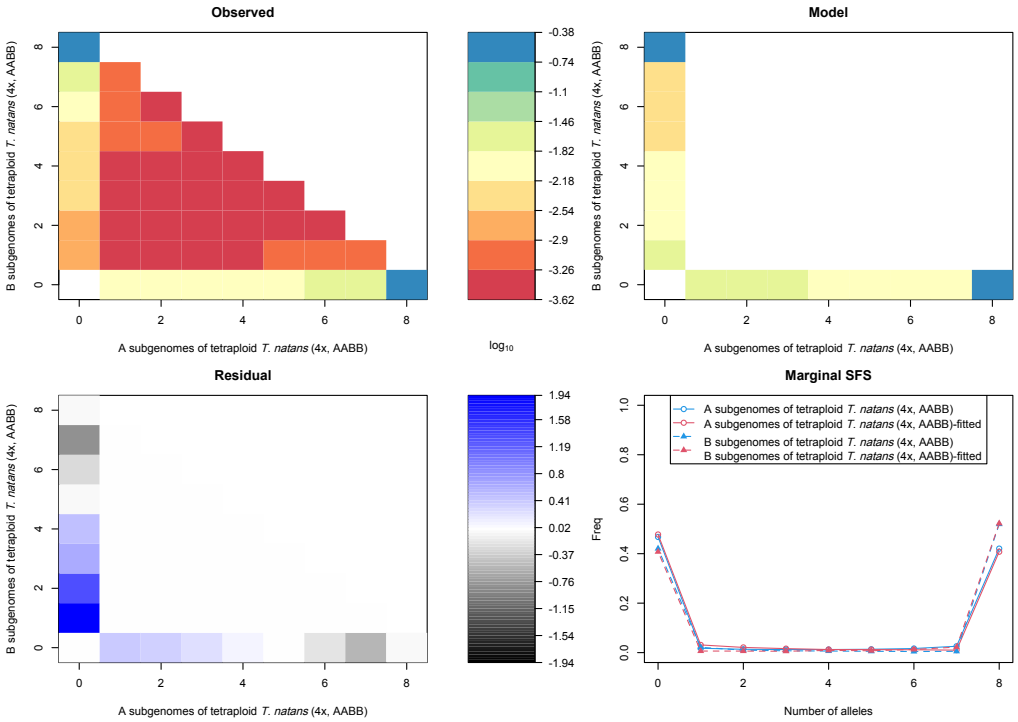



**Figure S7 (continued)**

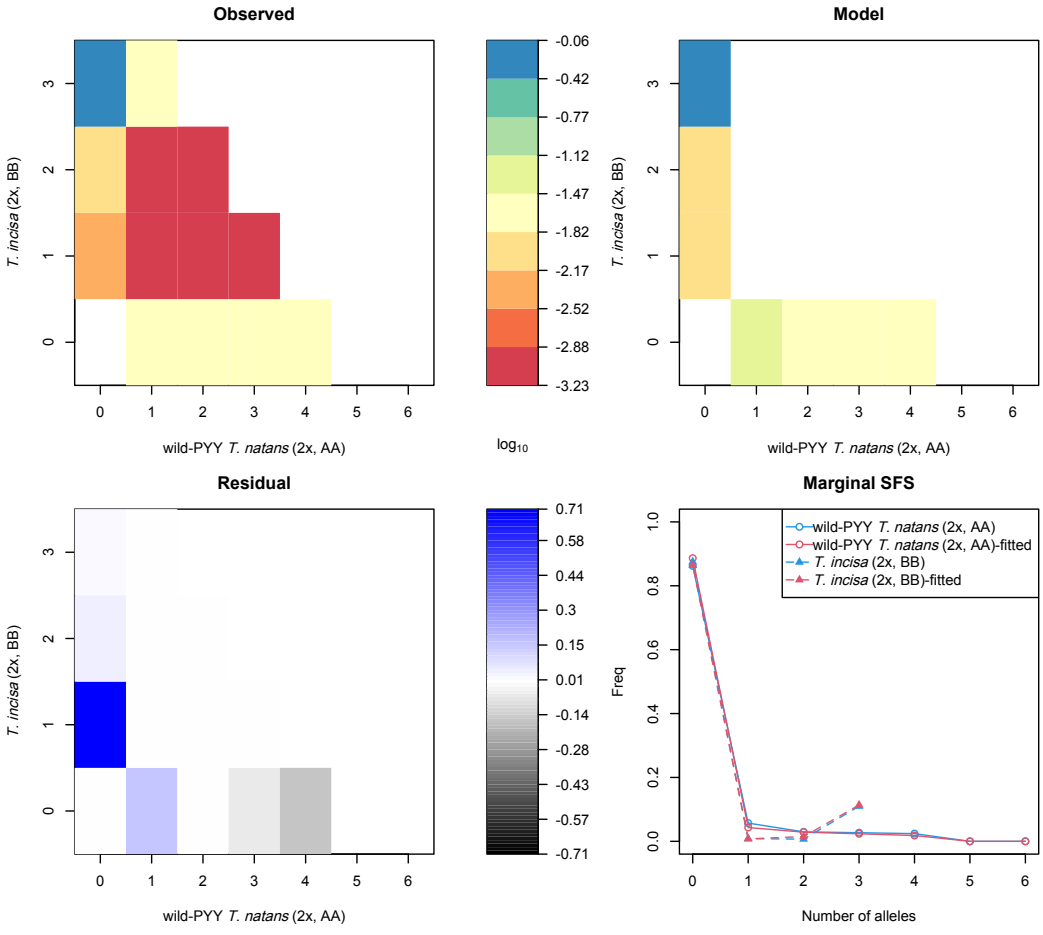



**Figure S7 (continued)**

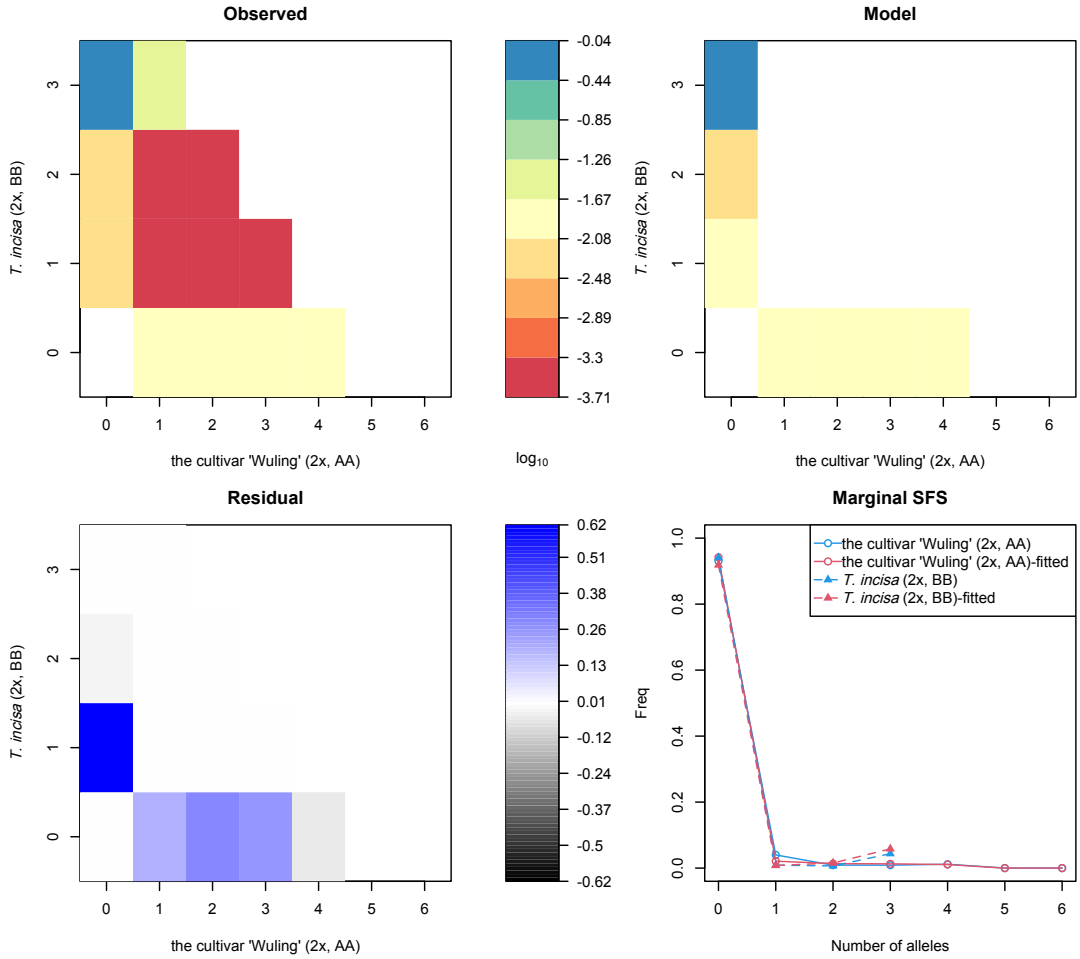



**Figure S7 (continued)**

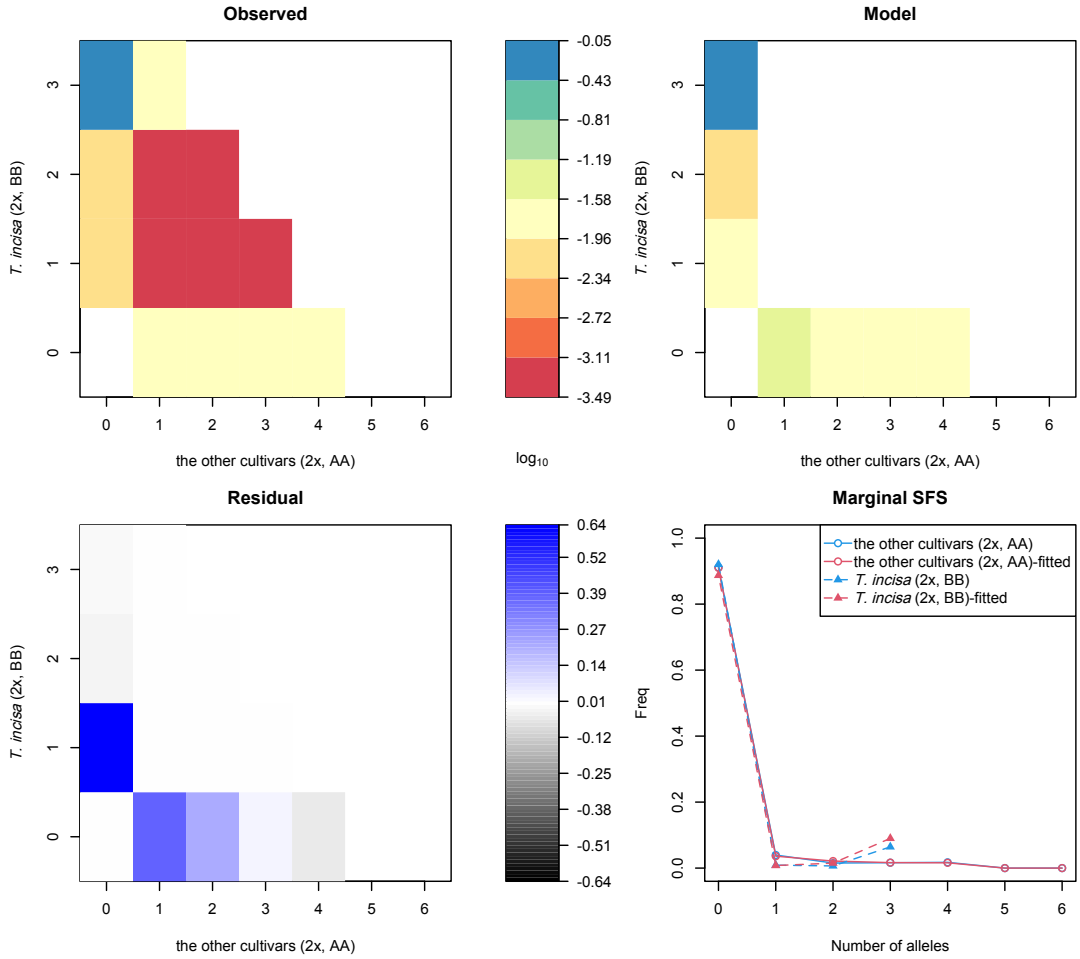



**Figure S7 (continued)**

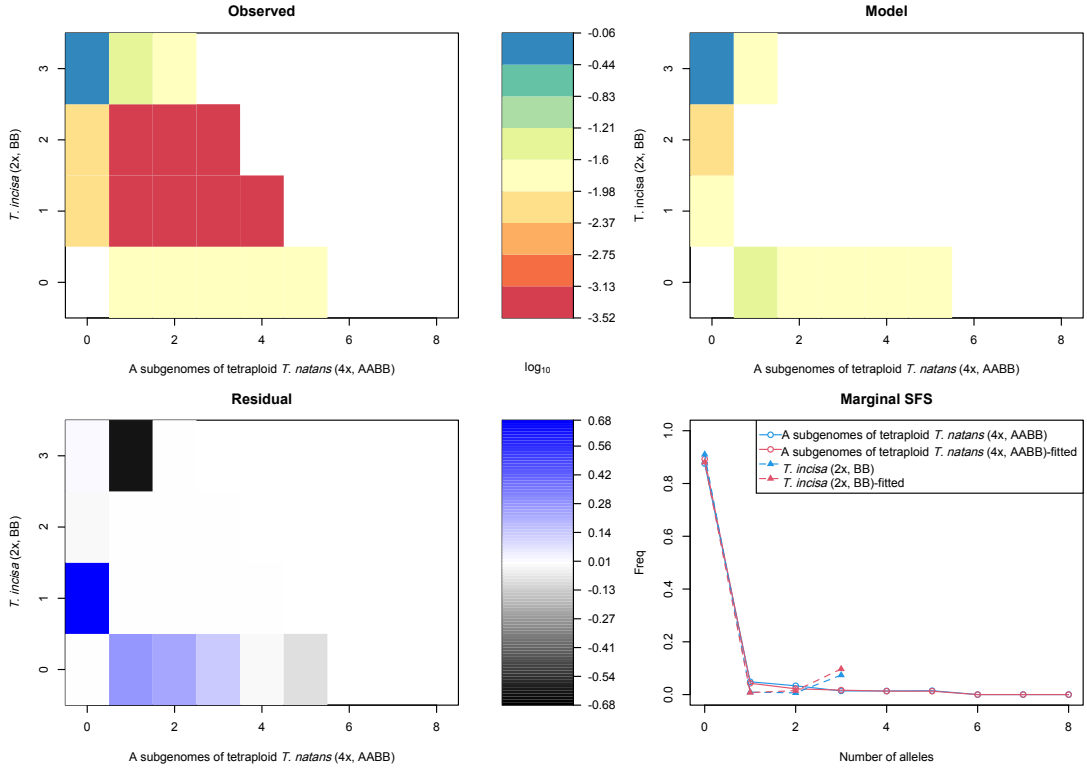



**Figure S7 (continued)**

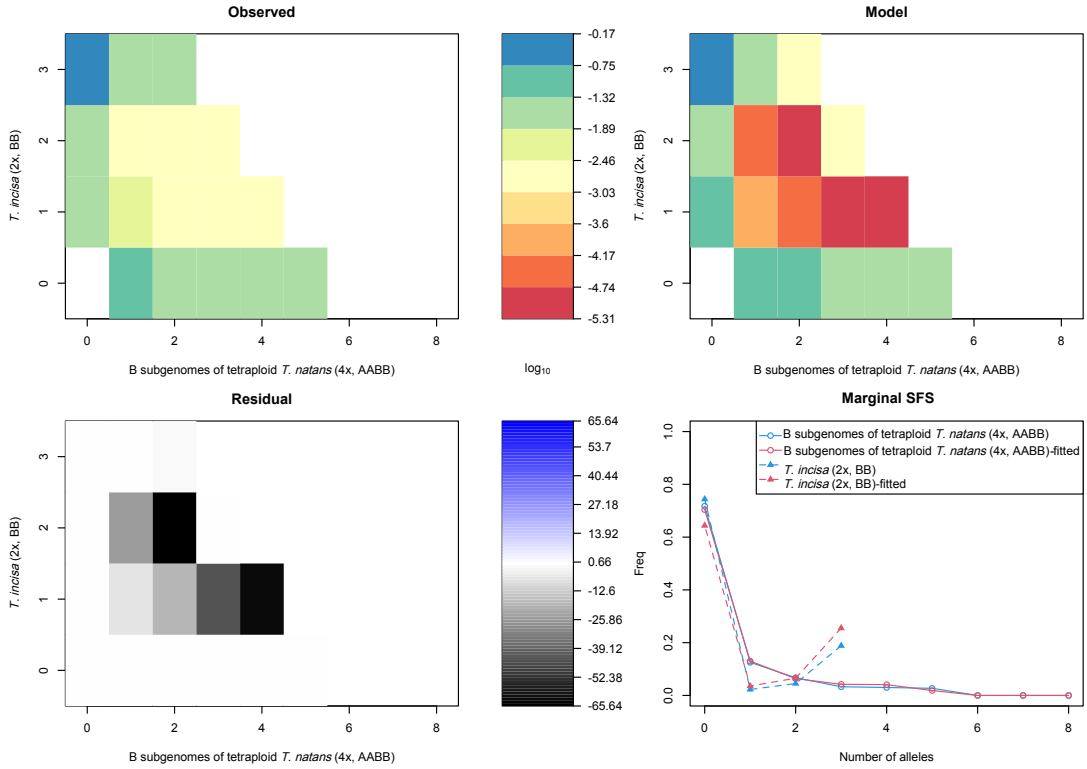



**Figure S7 (continued)**

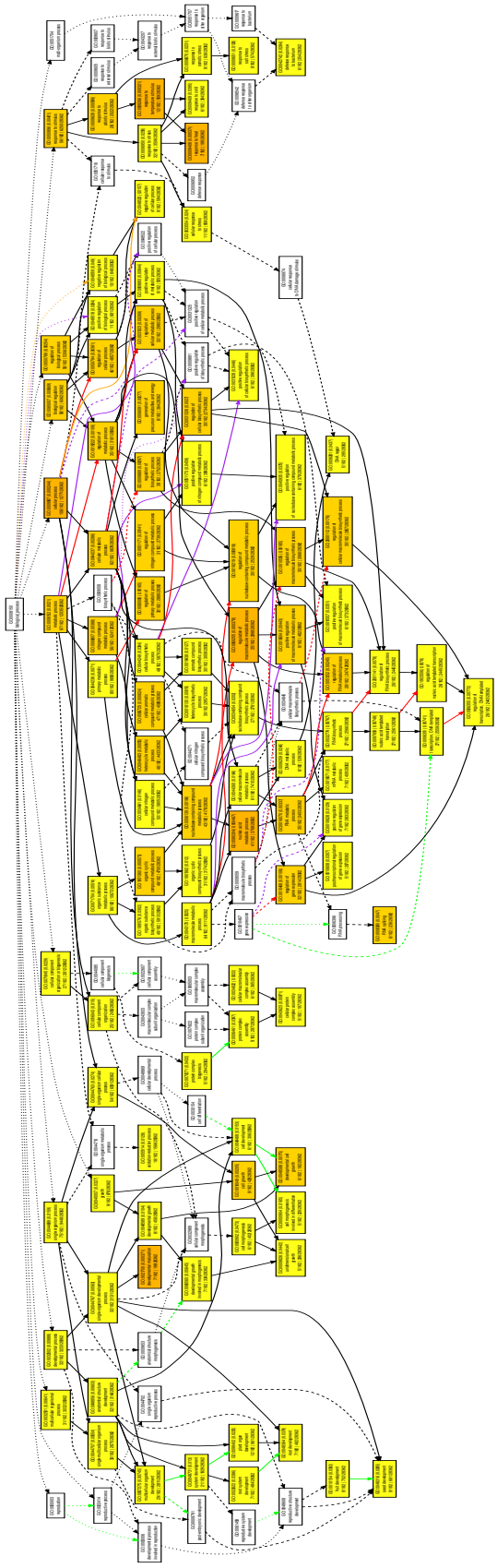


**Figure S8** Significantly enriched biological process GO categories of 205 candidate genes under positive selection.

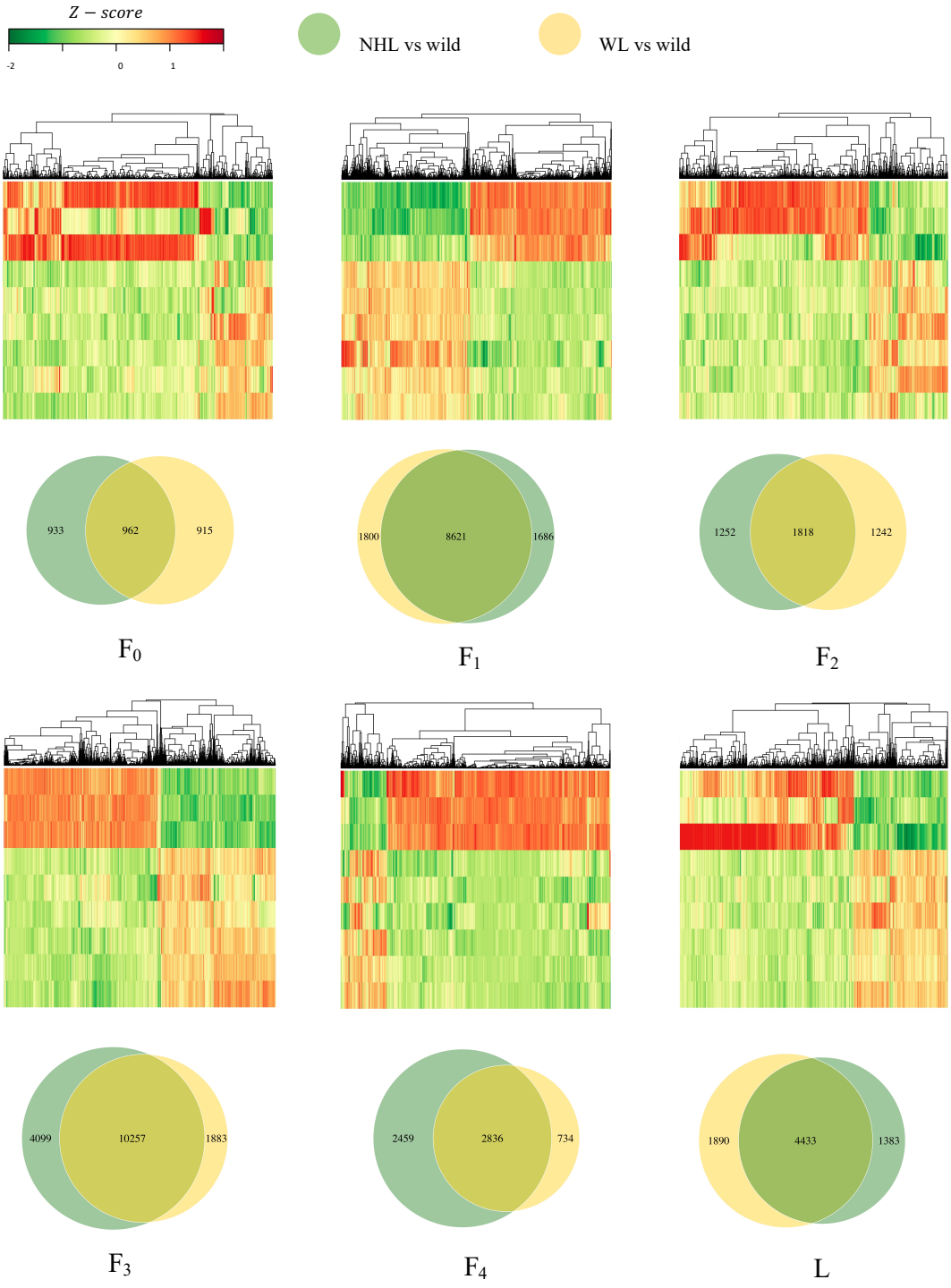
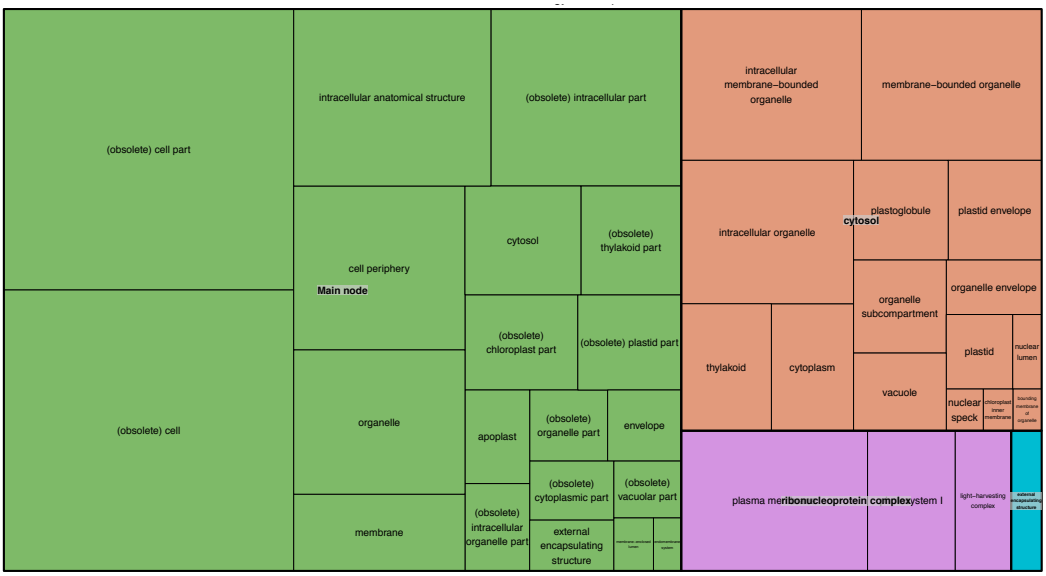



**Figure S9** Heatmaps and Venn diagrams of common differentially expressed (up- or downregulated) genes (DEGs) shared by the two cultivar–wild comparisons (i.e. 'Wuling' vs. wild and 'Nanhuling' vs. wild). $F_0$: flower bud, $F_1$: fertilized flower, $F_2$: fruit before sepal abscission, $F_3$: fruit after sepal abscission, $F_4$: juvenile fruit, L: leaf.

a




b

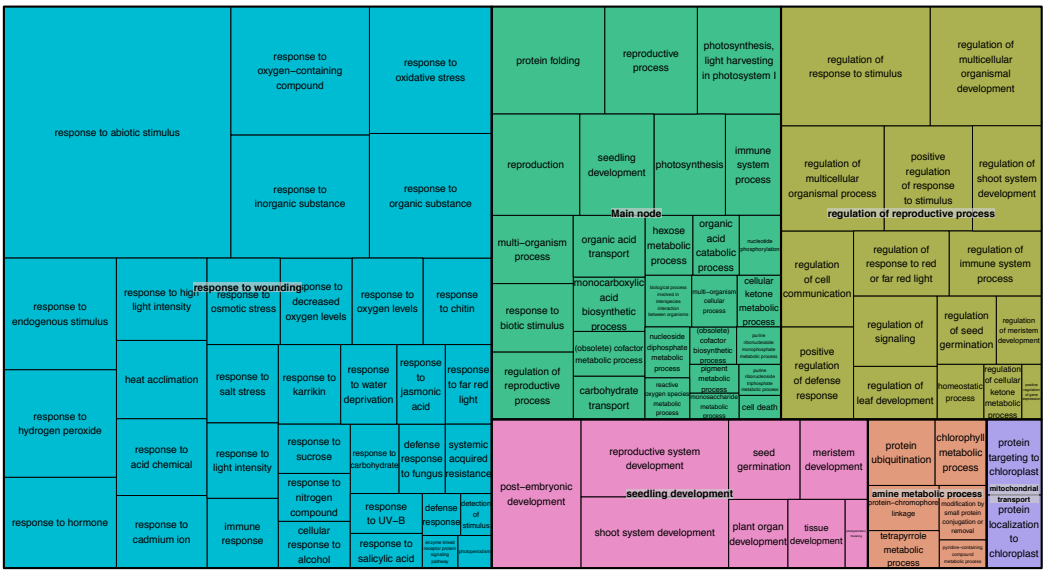
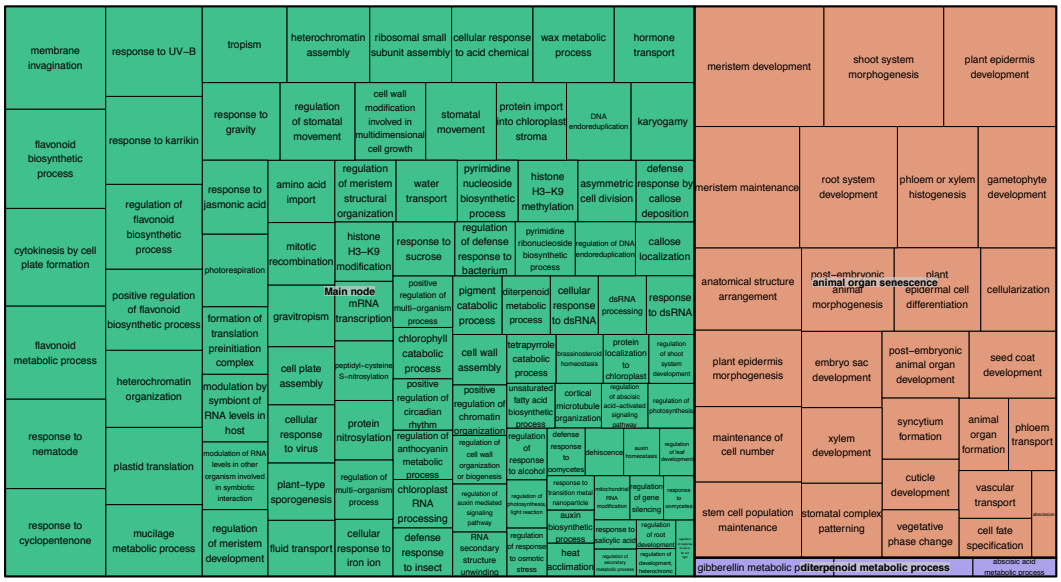



**Figure S10** Significantly enriched GO categories under biological process of differentially expressed (up- or downregulated) genes (DEGs) shared by the two cultivar–wild comparisons (i.e. ‘Wuling’ vs. wild and ‘Nanhuling’ vs. wild). a, c, e, g, i, upregulated genes at developmental stages, $F_0$ (flower bud), $F_1$ (fertilized flower), $F_2$ (fruit before sepal abscission), $F_3$ (fruit after sepal abscission) and $F_4$ (juvenile fruit), and in leaf tissues, respectively. b, d, g, h, k, downregulated genes at $F_0$, $F_1$, $F_2$, $F_3$ and $F_4$ stages and in leaf tissues, respectively.

c




d

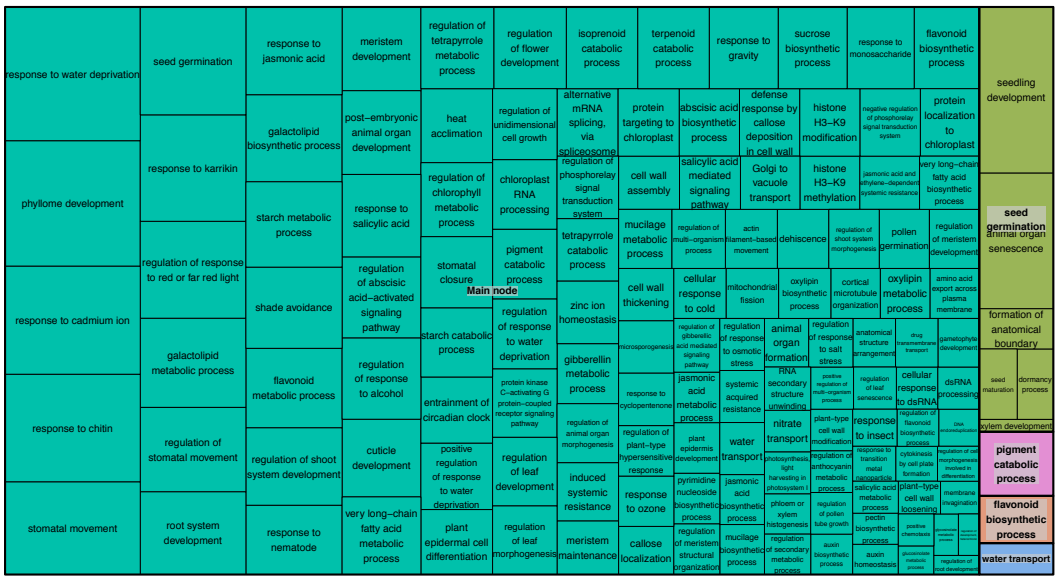
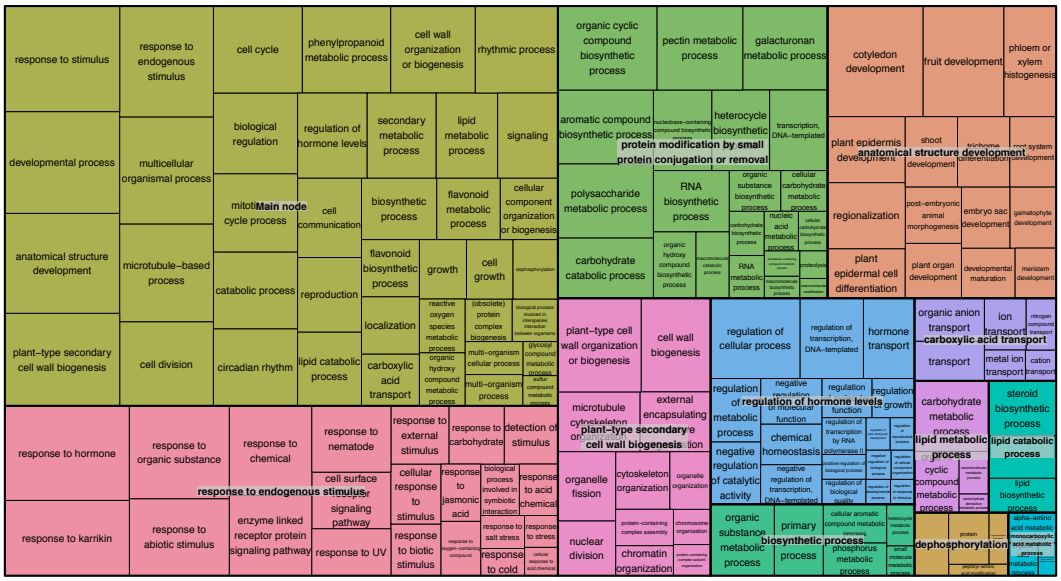



**Figure S10 (continued)**

e




f

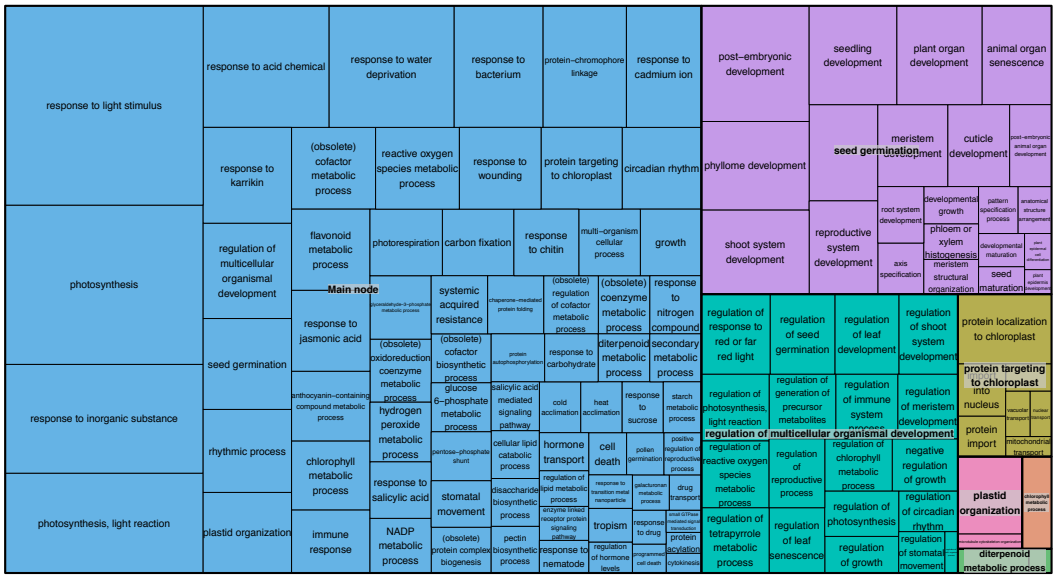
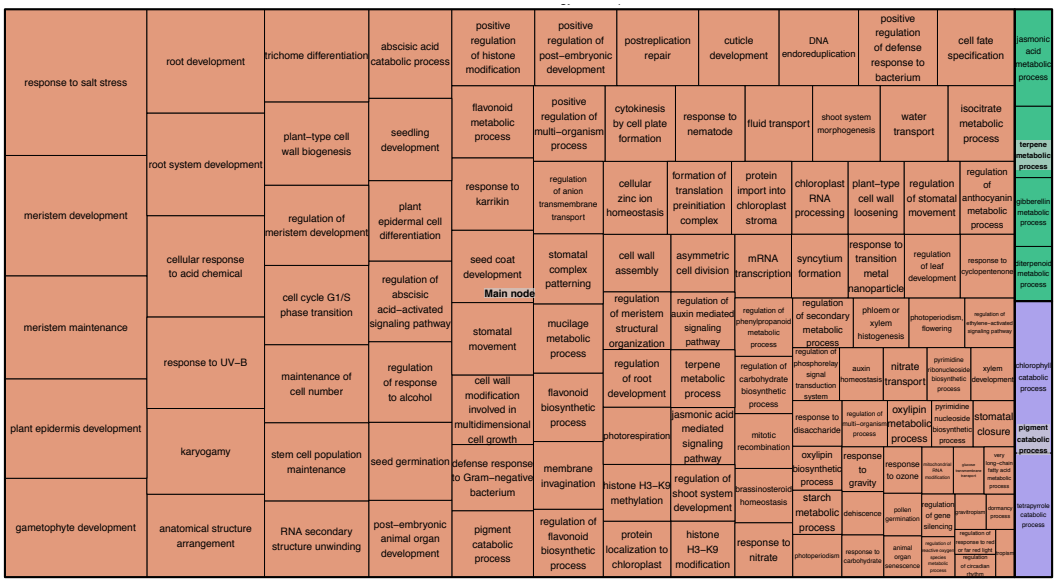



**Figure S10 (continued)**

g




h

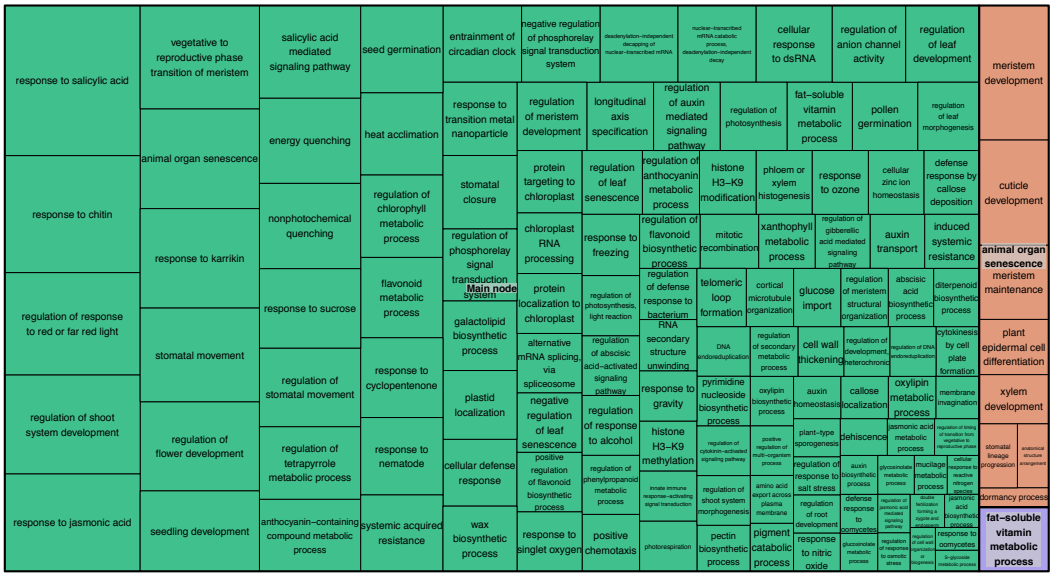
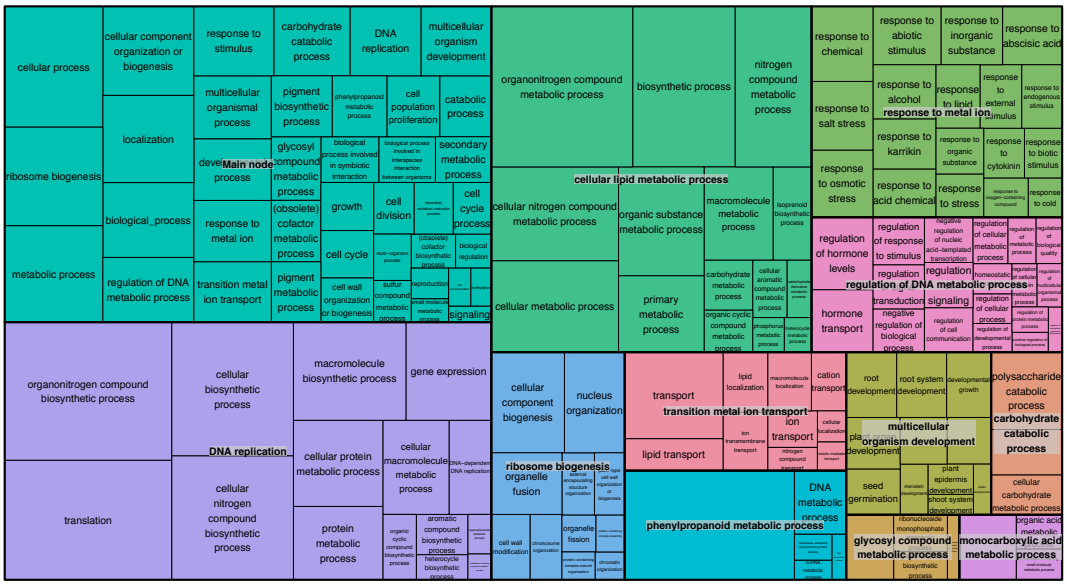



**Figure S10 (continued)**

i




j

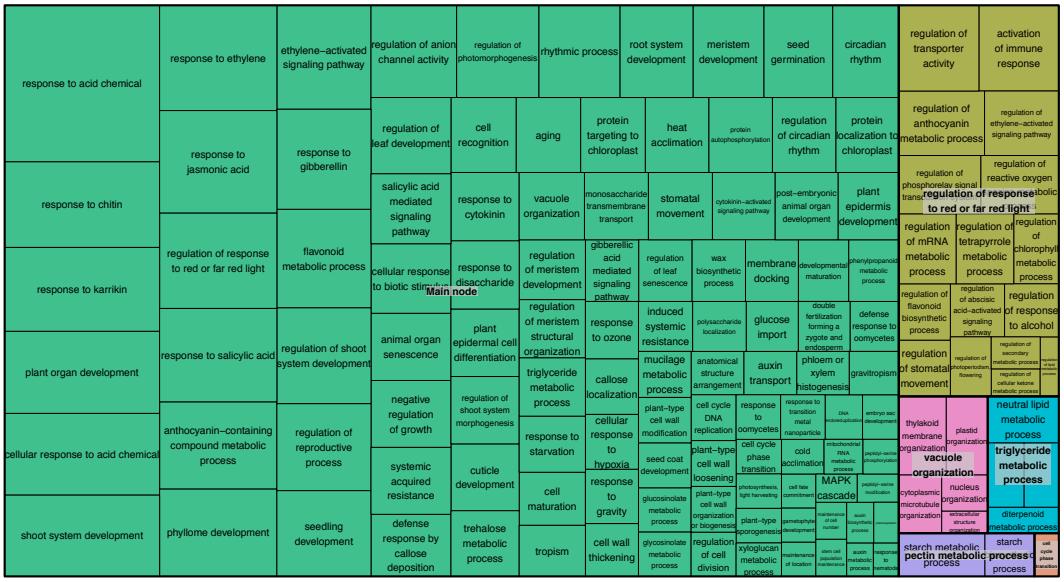
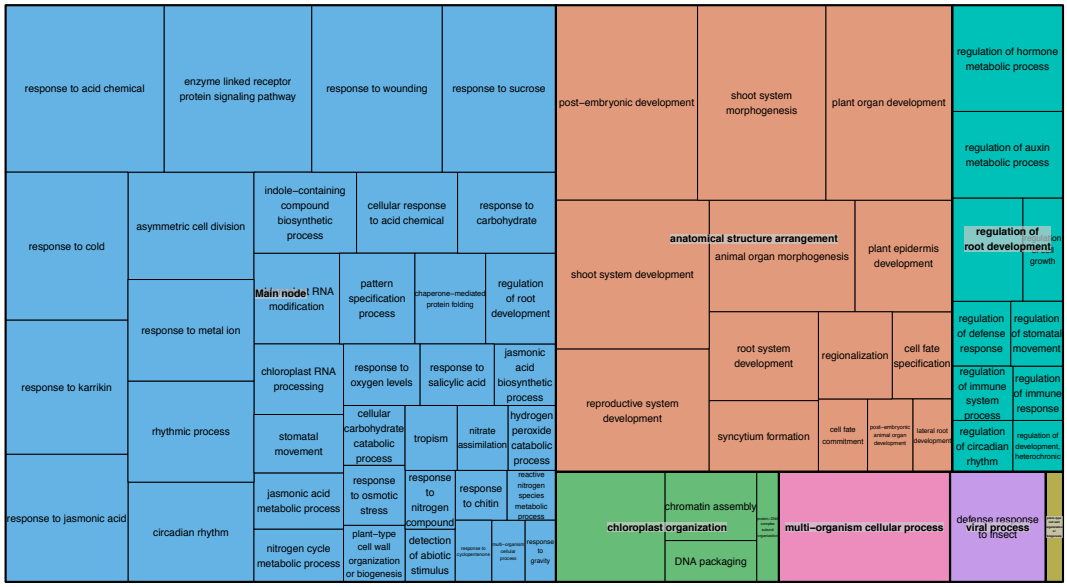



**Figure S10 (continued)**

k




l

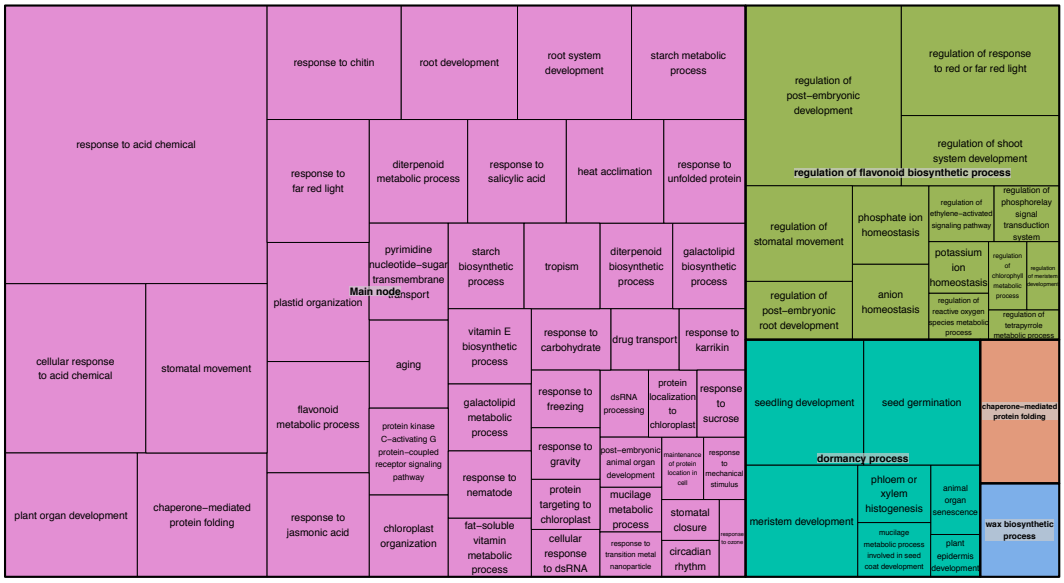
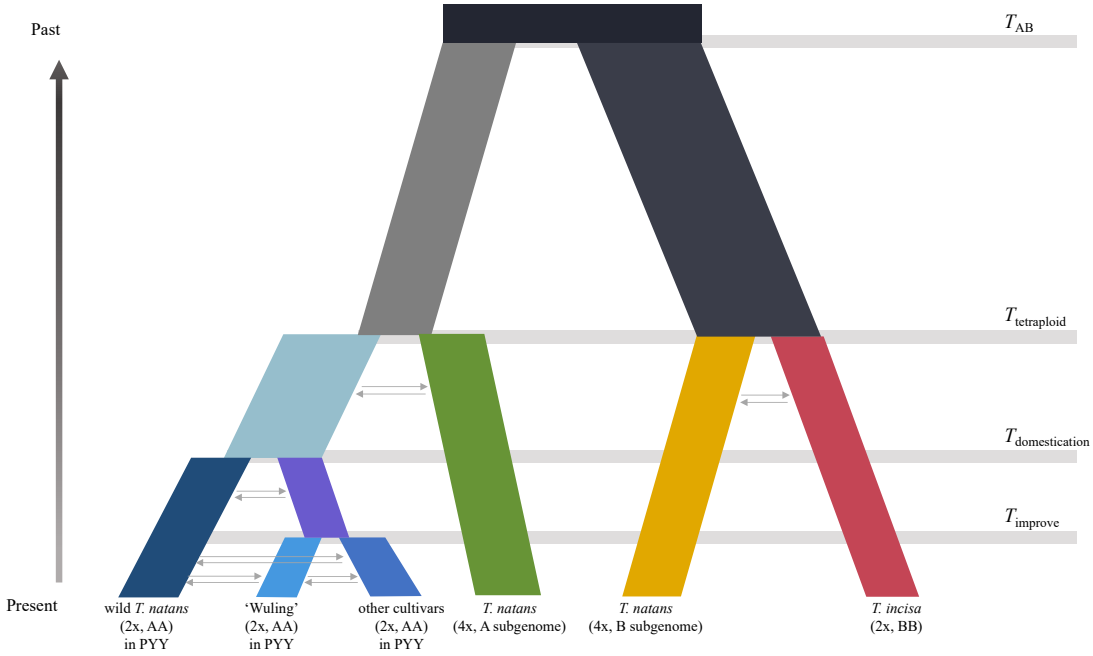
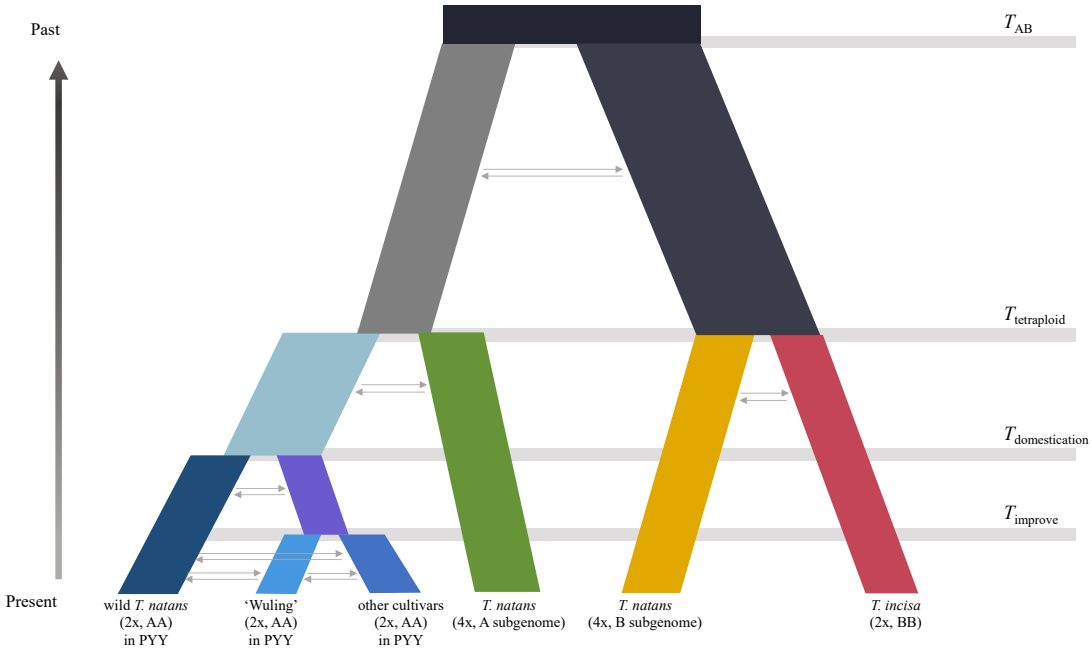



**Figure S10 (continued)**

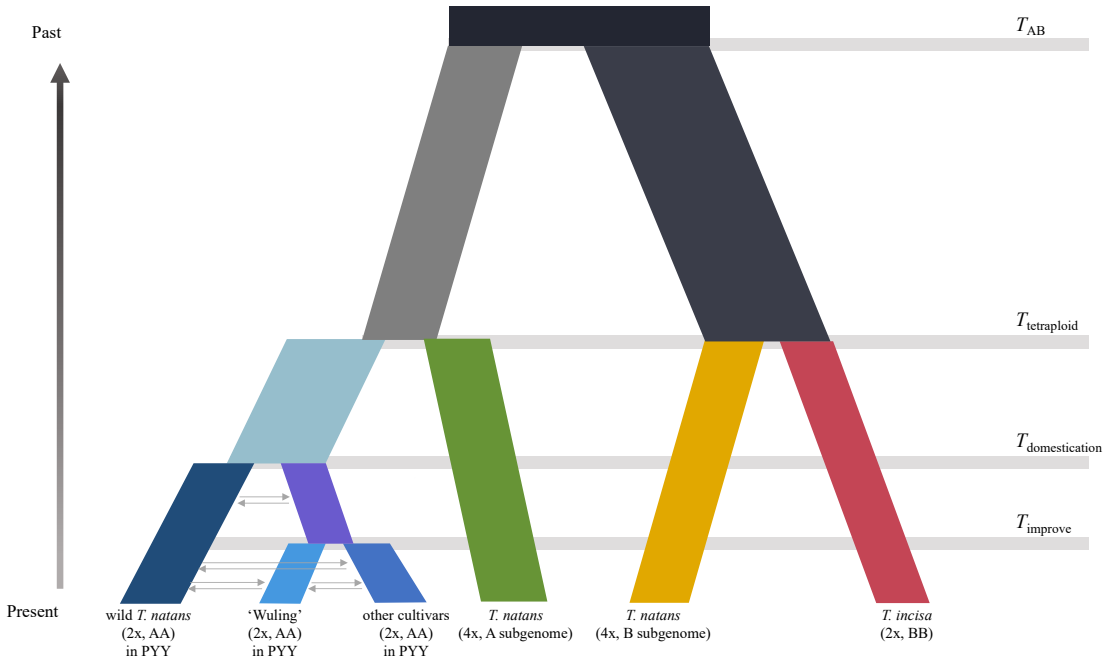



Model I

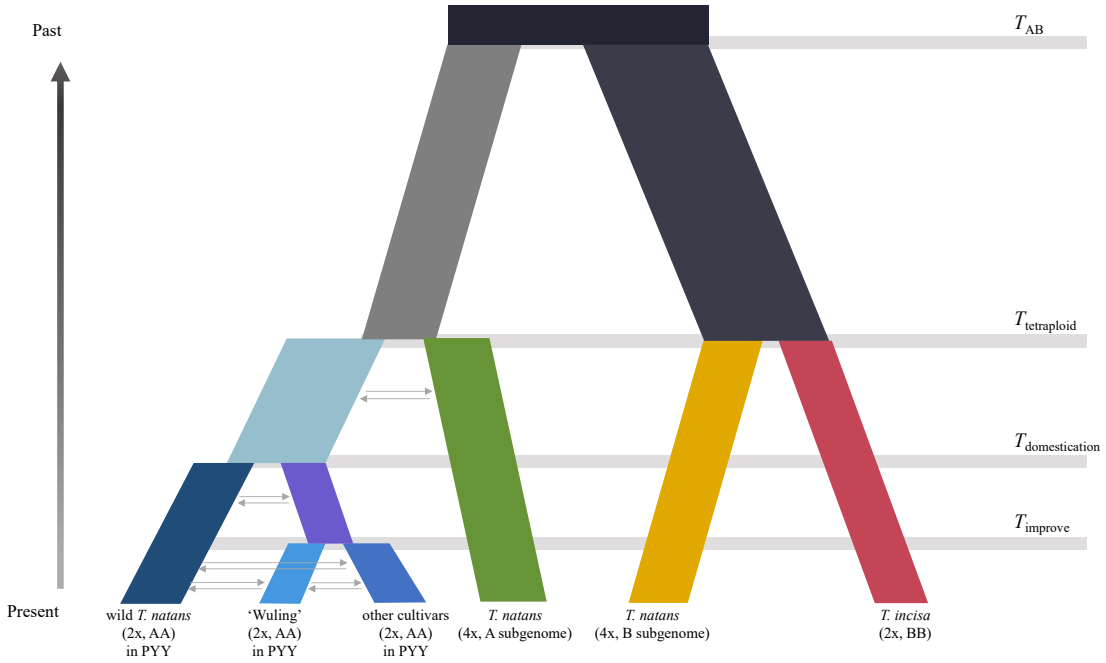



Model II

**Figure S11** Test of alternative demographic models, with different sets of migration directions using FASTSIMCOAL2. Parameter abbreviations: $T_{improve}$, time of improvement event for *T. natans* cultivar 'Wuling'; $T_{domestication}$, time of domestication event for all cultivars of *T. natans*; $T_{tetraploid}$, time of origin of allotetraploid *T. natans* (4x, AABB); $T_{AB}$, divergence time of diploid *T. natans* and *T. incisa*.




Model III




Model IV

**Figure S11 (continued).**

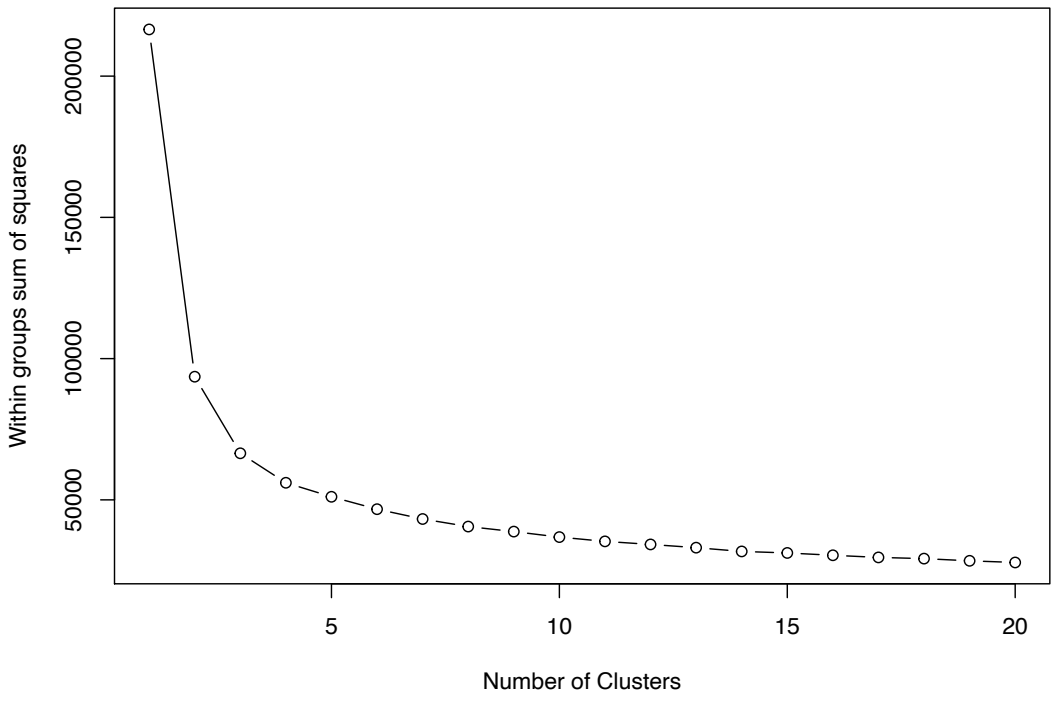



**Figure S12** The relationship between the sum of squared error (SSE) and the number of clusters. The number of clusters was set from 1 to 20.