

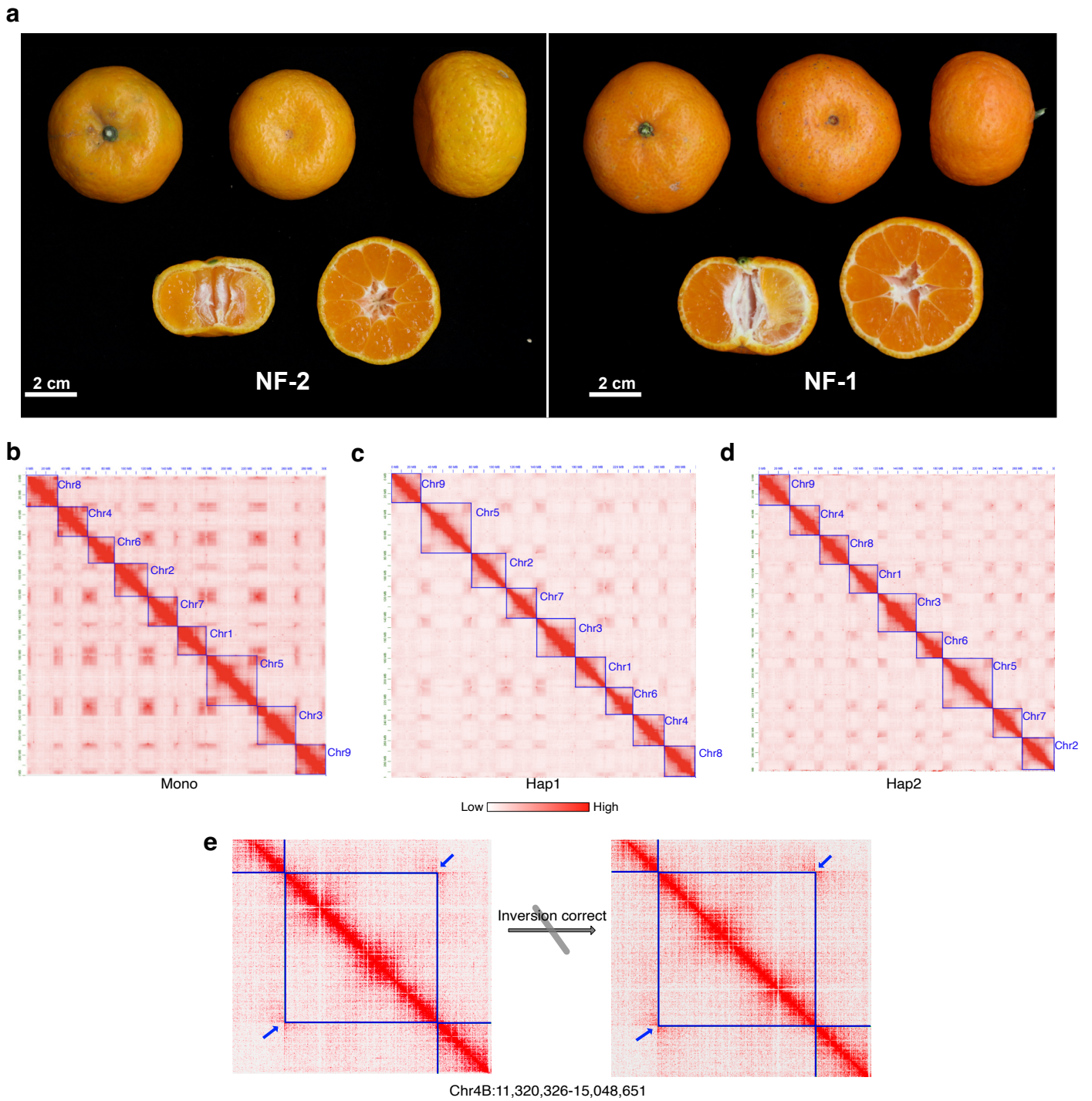


Fig. S1. The NFMJ_120y fruit phenotype and heat maps of Mono, Hap1 and Hap2 showing the distribution of Hi-C interaction signals in a 500 kb resolution.

a. The NF-1 and NF-2 fruit phenotype.

(b-d): Genome-wide Hi-C interaction heat map.

(e) The juicerbox was used to correct the inversion at Chr4B : 11,320,326-15,048,651, but the Hi-C signal at both ends of the chromatin was weakened after correction, indicating that the position does not need to be corrected, and the inversion really exists. High densities of interactions are indicated in red.

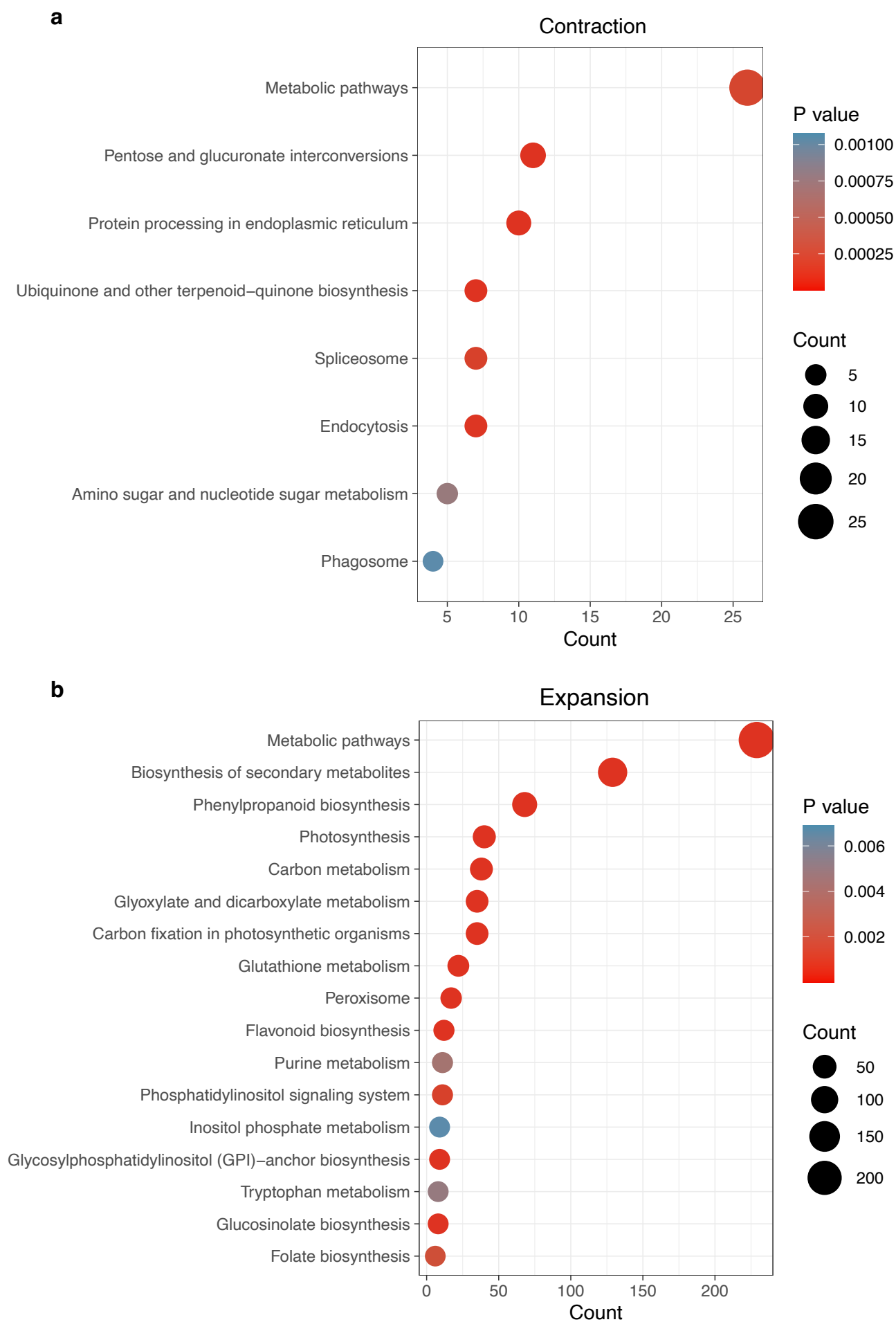


Fig. S2. KEGG pathway analysis of expanding and contracting gene families.

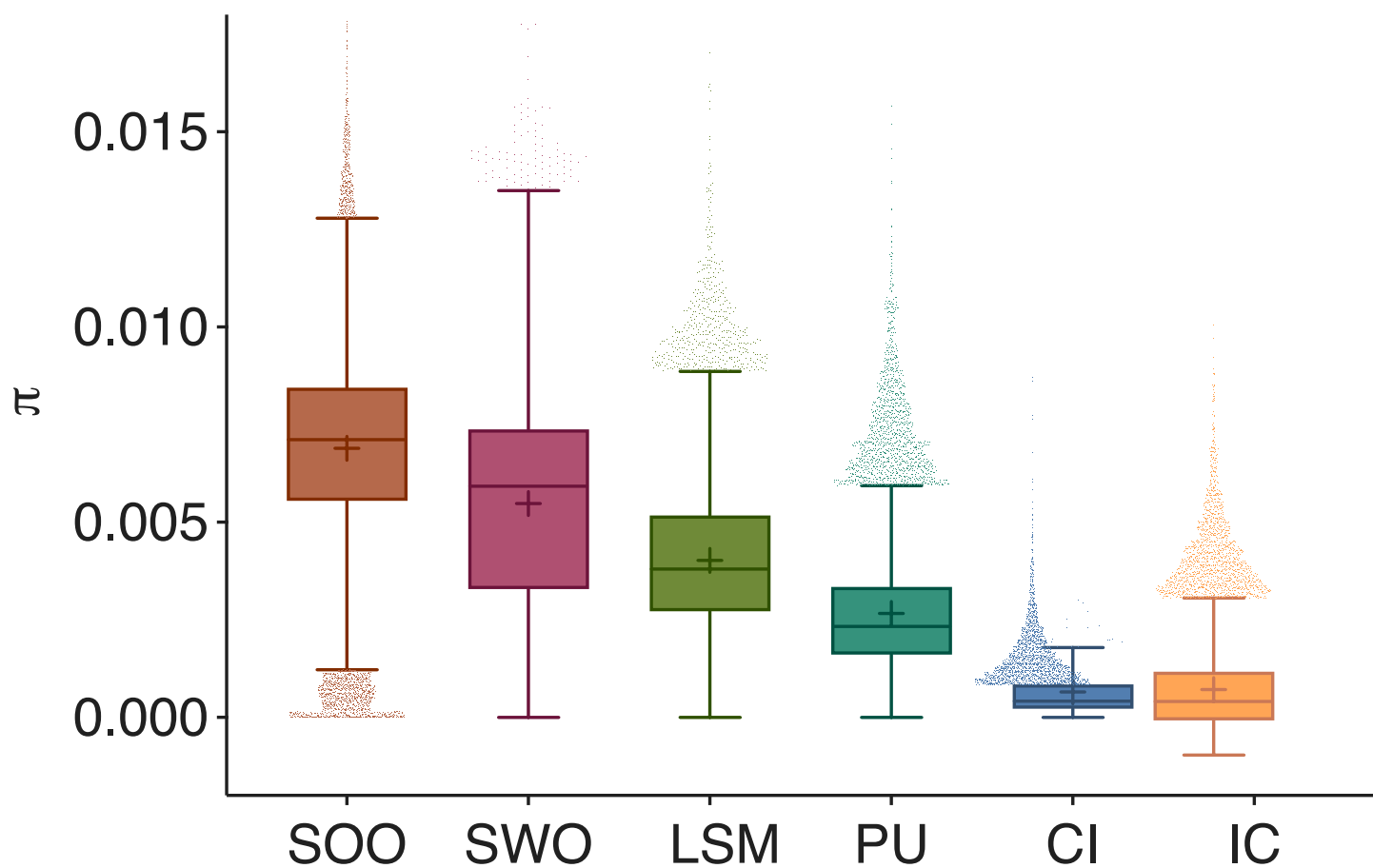


Fig. S3. Nucleotide diversity of different citrus species

SOO: Sour orange, SWO: Sweet orange, LSM: Loose-skin mandarin, PU: Pummelo, CI: Citron, IC: Ichang papeda

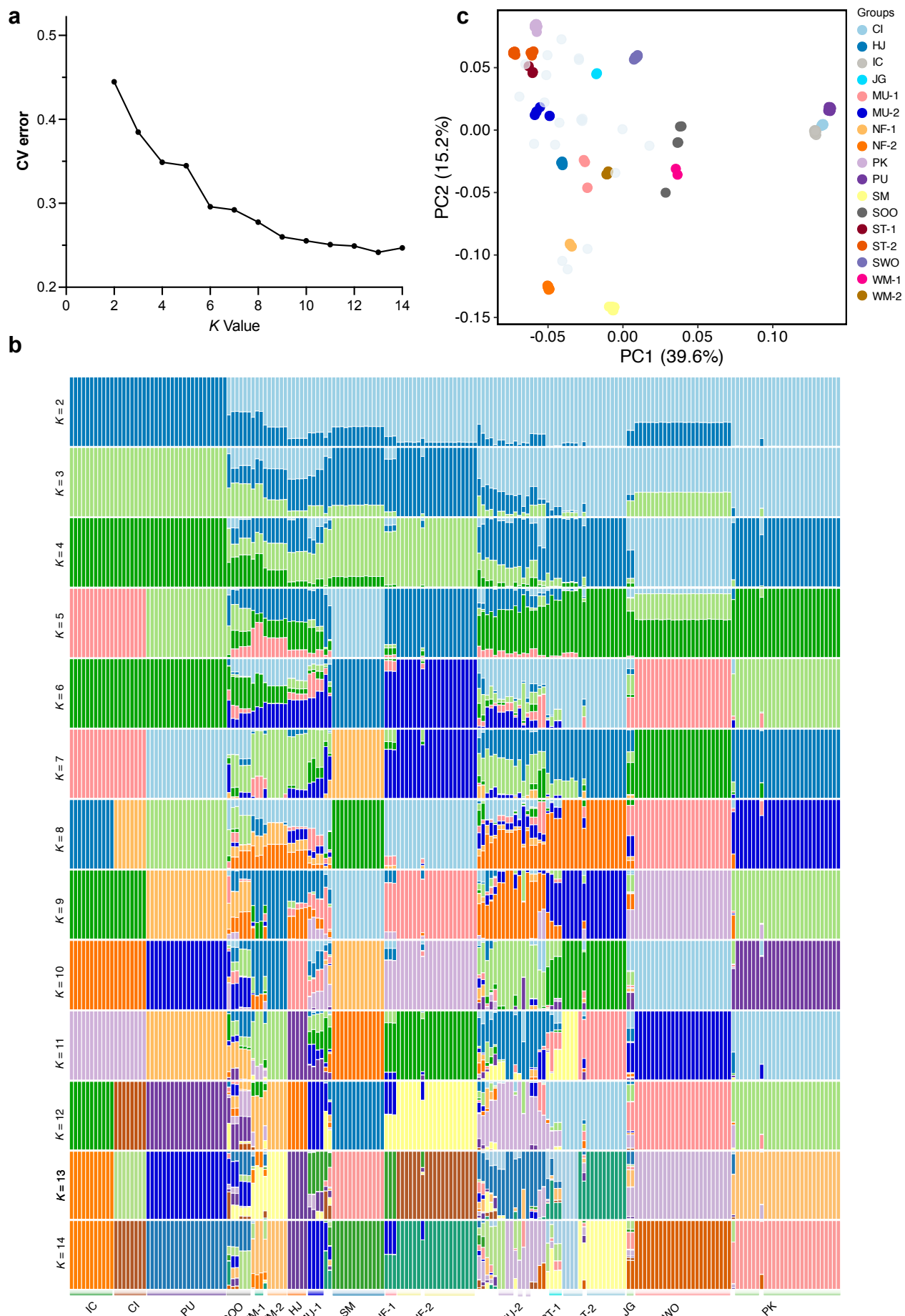


Fig. S4.The population structure and ancestry composition estimated by ADMIXTURE in mandarin population.

a Cross-validation error plot for the unsupervised ADMIXTURE analysis.

b. ADMIXTURE clustering of core accessions from K=2 to 14.

c. Principal component analysis (PCA) was performed after 12 groups were divided.

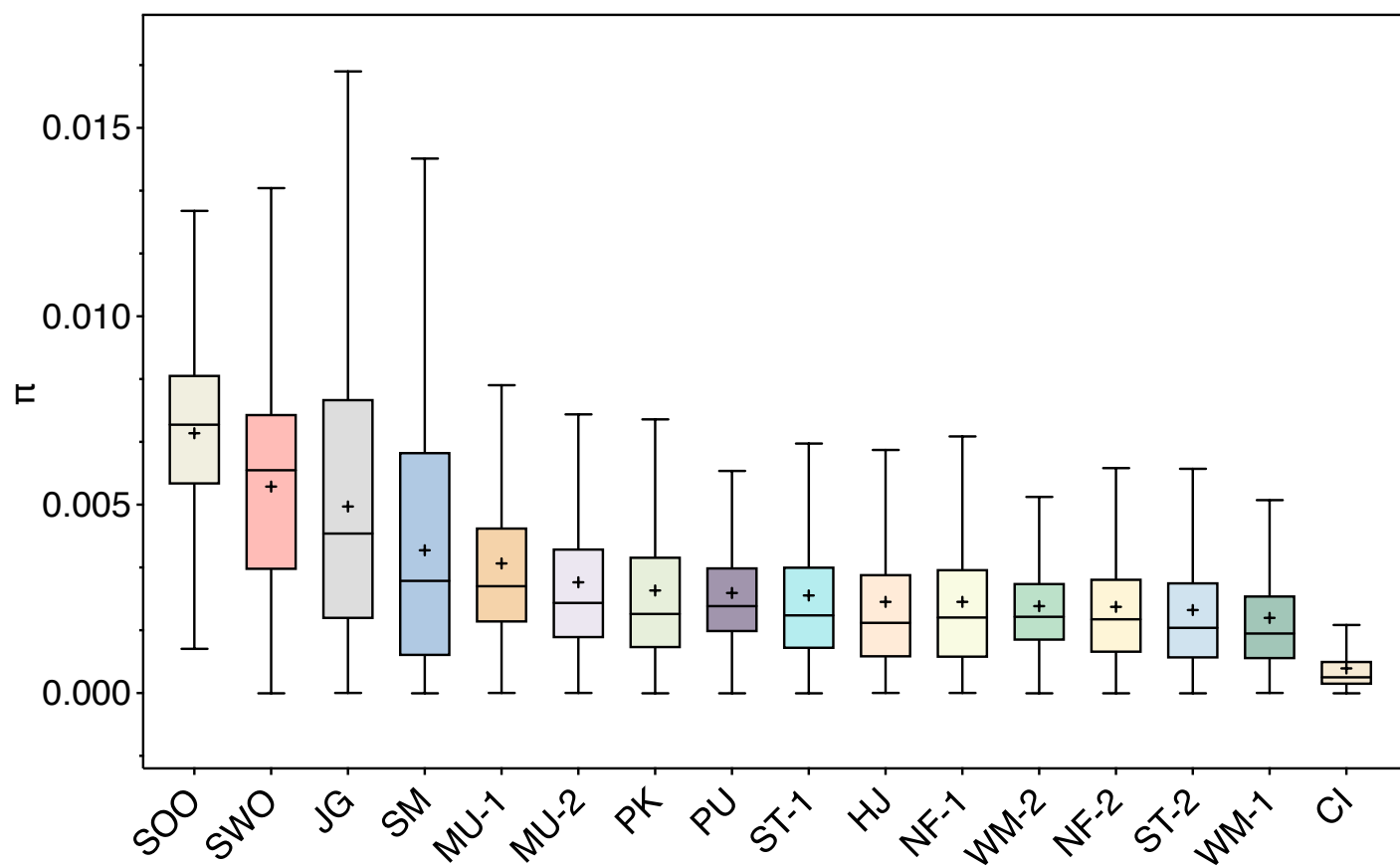


Fig. S5. Nucleotide diversity (π) of 16 groups.



Fig. S6. Distribution of LSMs in southern China.

YZ: Yangtze river group; PR: Pearl river group; PK: Ponkan; WM: Wild mandarin

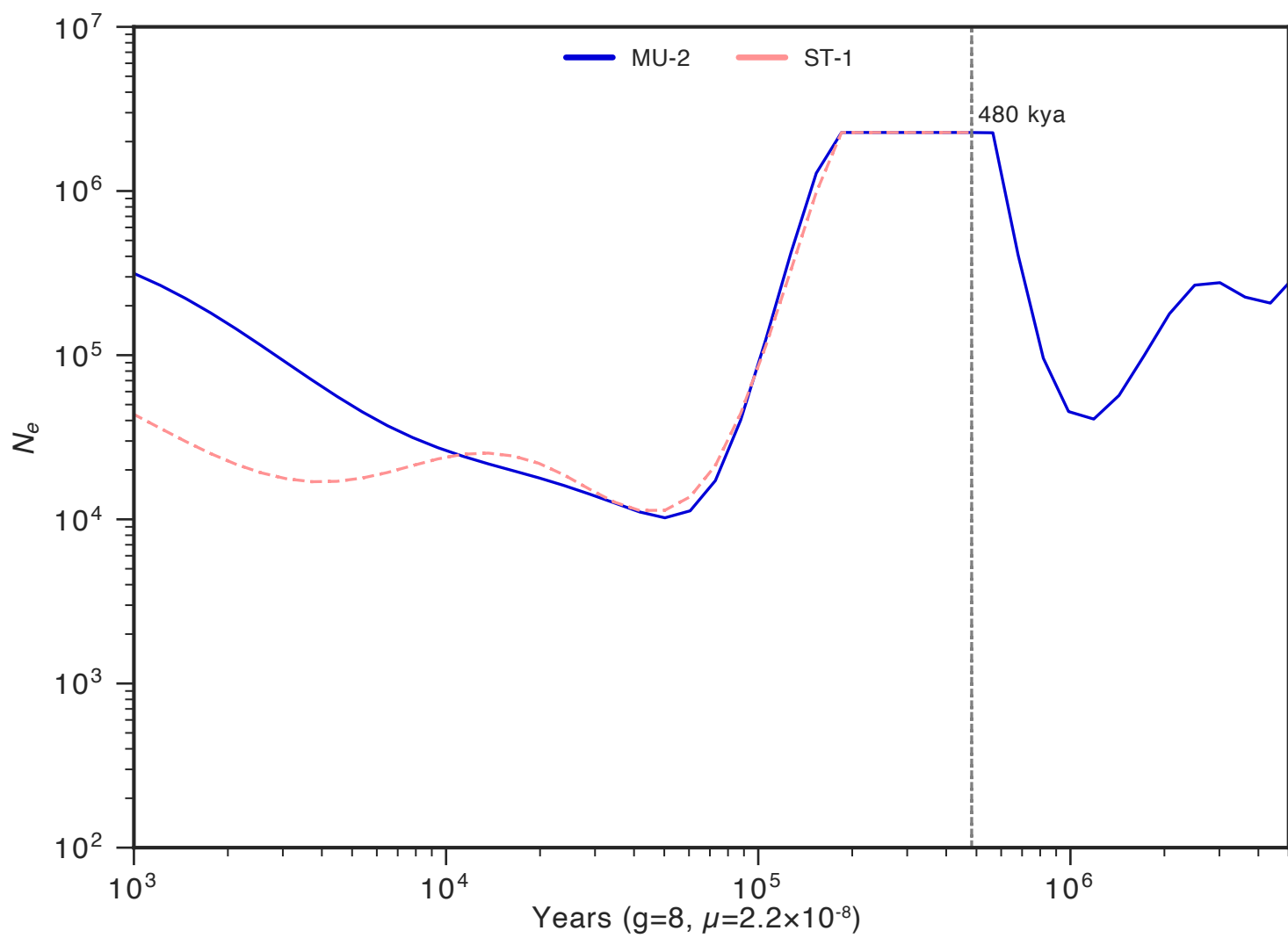


Fig. S7. Effective population size and split time between MU-2 and ST-1 inferred by SMC++.

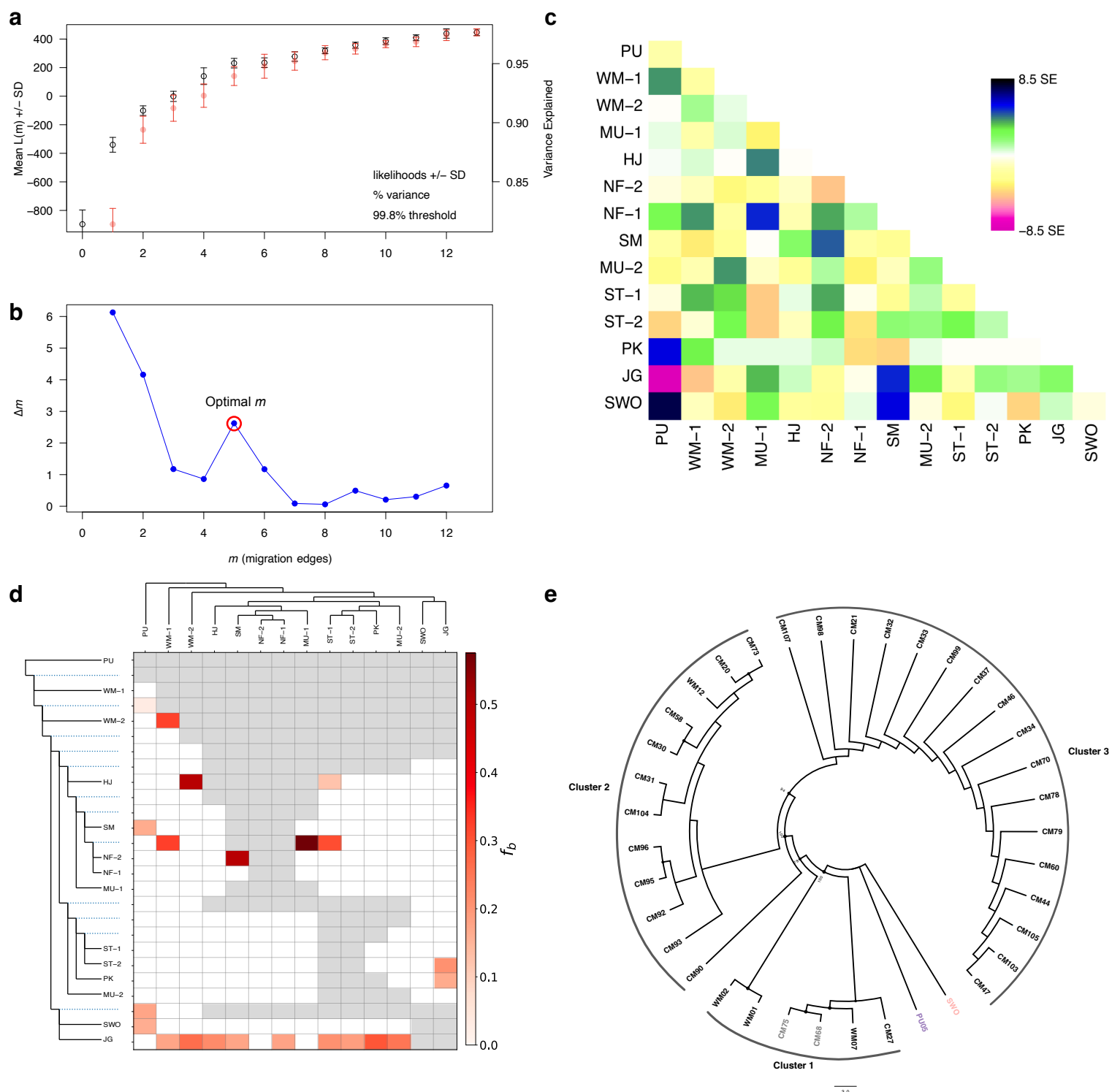


Fig. S8. Analysis of introgression in different citrus groups

(a) The mean and standard deviation (SD) across 12 iterations for the composite likelihood $L(m)$ (left axis, black circles) and proportion of variance explained (right axis, red dots).

(b) The second-order rate of change (Δm) across values of m , the optimal m number shown by the red circle.

(c) The residual fit plots from the maximum likelihood tree in Supplementary Figure11D.

(d) Branch introgression (fb) statistic. Values of the branch introgression (fb) statistic represent the potential signal between groups.

(e) Phylogenetic tree constructed by chloroplast genomes.

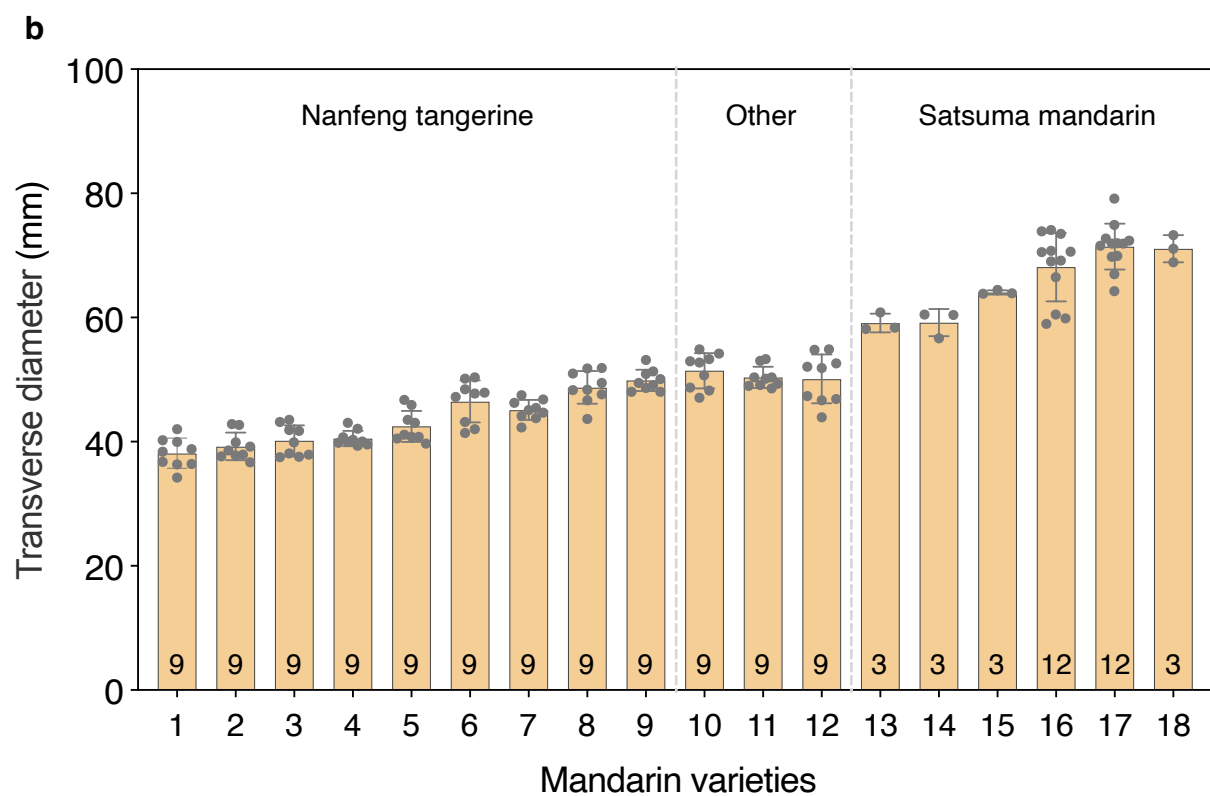
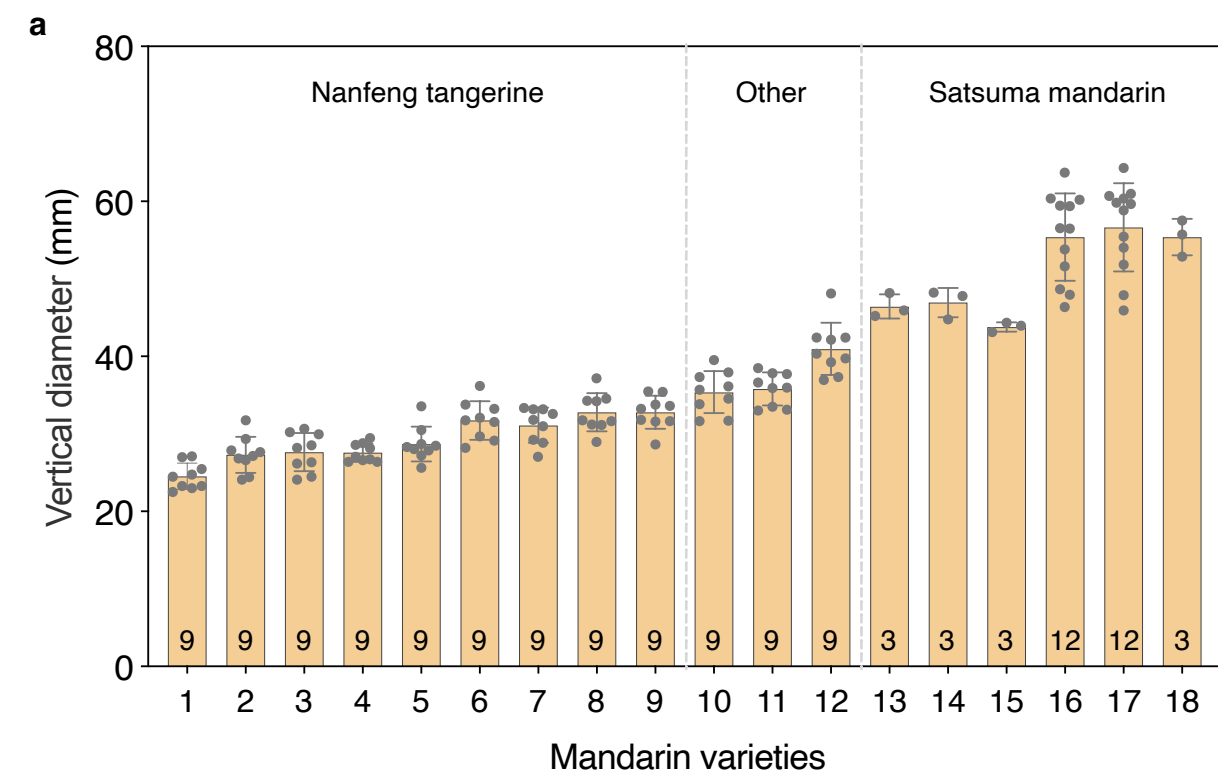


Fig. S9. Fruit vertical and transverse diameters of Nanfengmiju tangerine and Satsuma mandarin.

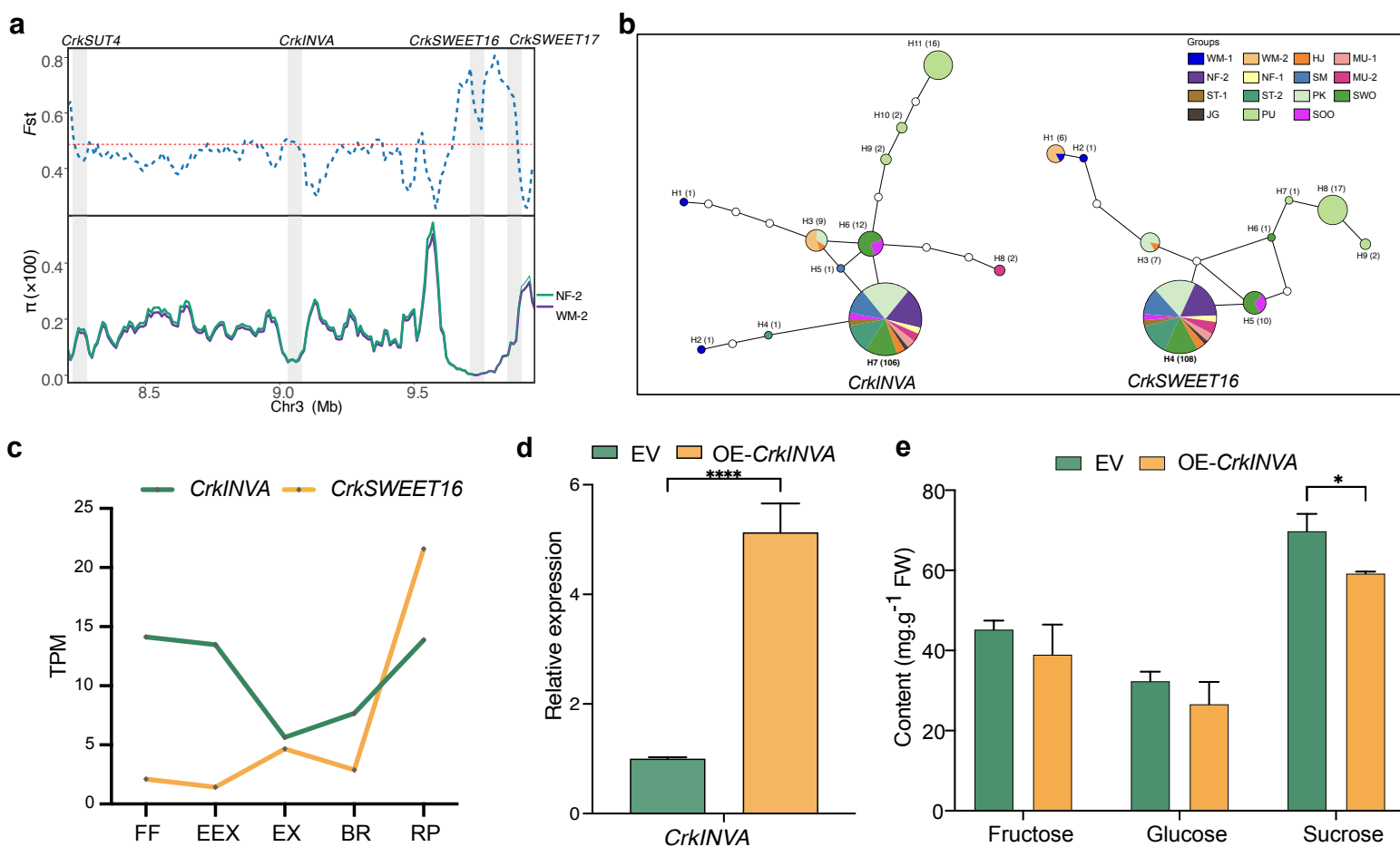


Fig. S10. The candidate selective gene of NF-2 vs. WM-2.

(a) Signals of artificial selection in the candidate gene of NF-2 vs. WM-2. Purple and green solid lines indicate π value statistics in WM-2 and NF-2, respectively. The blue dashed line indicates the F_{st} value between NF-2 and WM-2, while the red dashed line is the threshold of the top 10% F_{st} .

(b) The haplotype network analysis of *CrkINVA* and *CrkSWEET16*. Each circle represents one haplotype and the circle area is proportional to the frequency of each haplotype. Different colors represent different populations. Small uncoloured circles represent non-observed haplotypes; each line connecting haplotypes represents a single mutational change.

(c) Expression patterns of *CrkINVA* and *CrkSWEET16* of NF-2 vs. WM-2 in the JS during development of GYH. FF: the fruit-falling period, EEX, the Early expansion period.

(d) Relative expression of the *CrkINVA* in OE-*CrkINVA* lines (attached fruit). The asterisk indicates significance two-sided Student's t-test, The P value is 0.000004.

(e) Fruit soluble sugar content of OE-*CrkINVA* lines (attached fruit). The asterisk indicates significance two-sided Student's t-test, The P value for sucrose is 0.0135.

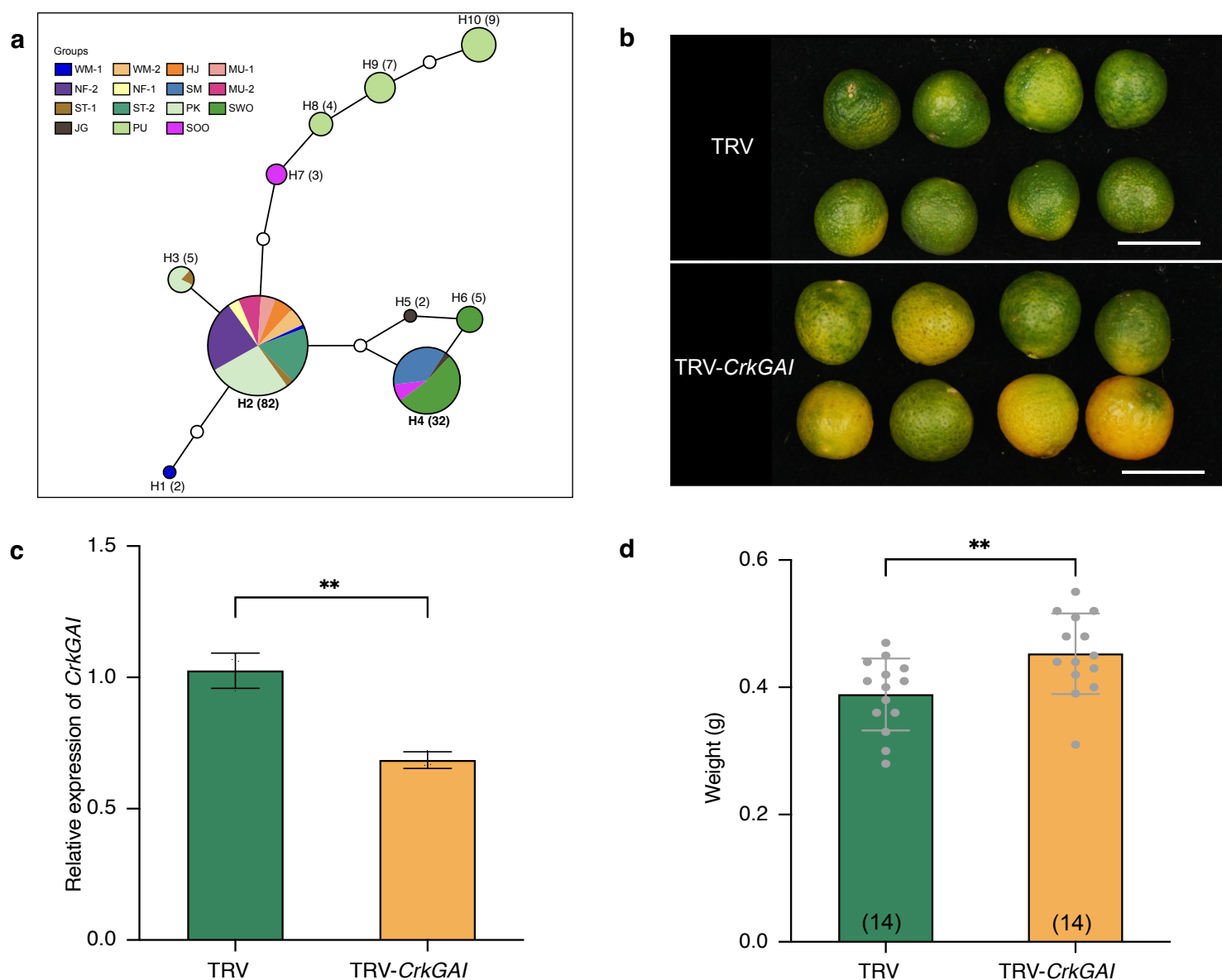


Fig. S11. The haplotype network analysis and transient silence of *CrkGAI* by VIGS in Hongkong kumquat fruits.

(a) The haplotype network analysis of *CrkGAI*. Each circle represents one haplotype and the circle area is proportional to the frequency of each haplotype. Different colors represent different populations. Small uncoloured circles represent non-observed haplotypes; each line connecting haplotypes represents a single mutational change.

(b) The fruits phenotype of TRV and TRV-*CrkGAI*.

(c) Relative expression of the *CrkGAI* in TRV and TRV-*CrkGAI*. The asterisk indicates significance with two-sided Student's t-test, the *P* values is 0.0014.

(d) The fruit weight of TRV and TRV-*CrkGAI*. The asterisk indicates significance with two-sided Student's t-test, the *P* value is 0.0088.

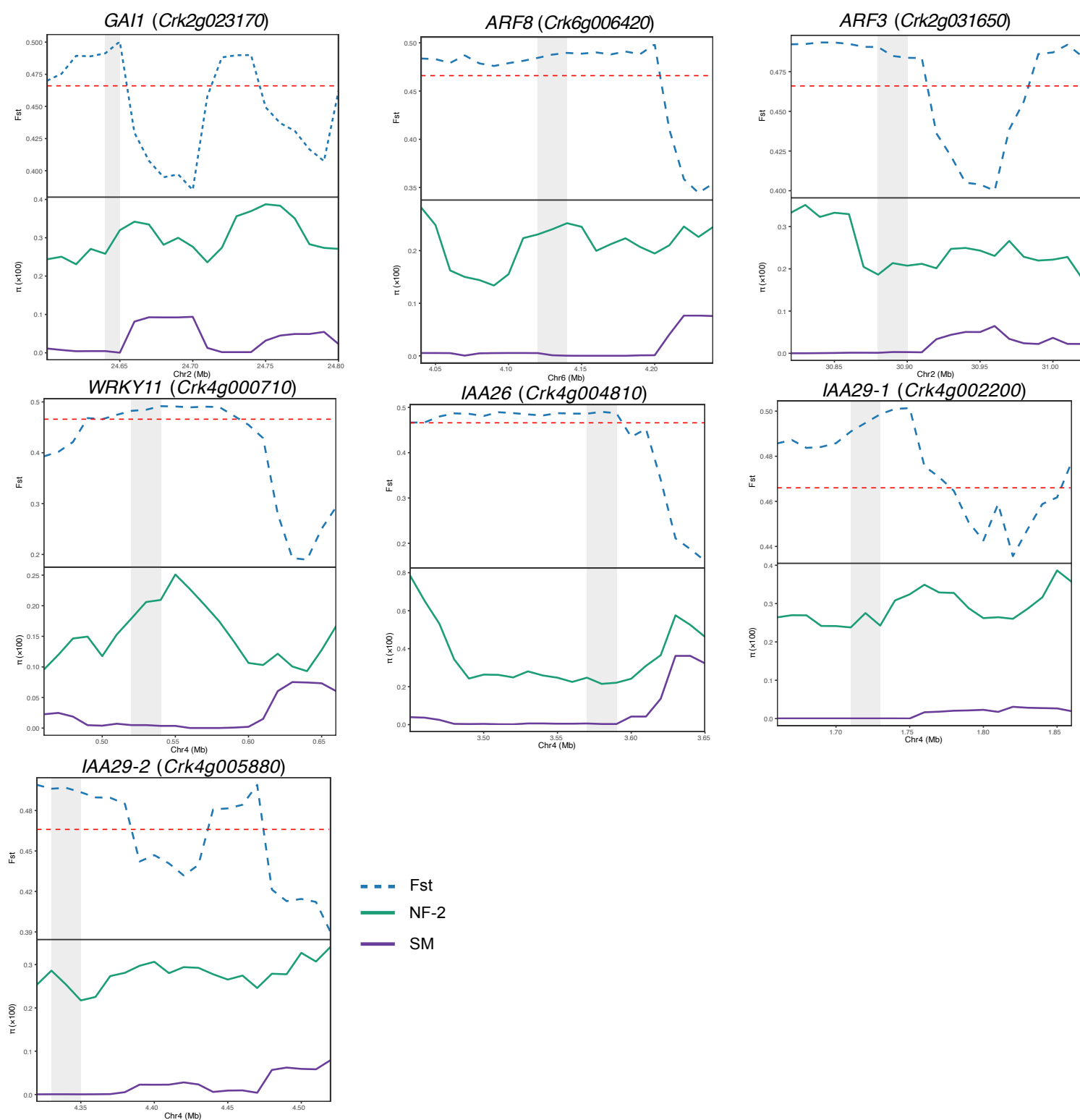


Fig. S12. Signals of artificial selection in the candidate genes of SM vs. NF-2. Purple and green solid lines indicate π value statistics in SM and NF-2, respectively. The blue dashed line indicates the F_{st} value between SM and NF-2, while the red dashed line is the threshold of the top 10% F_{st} .

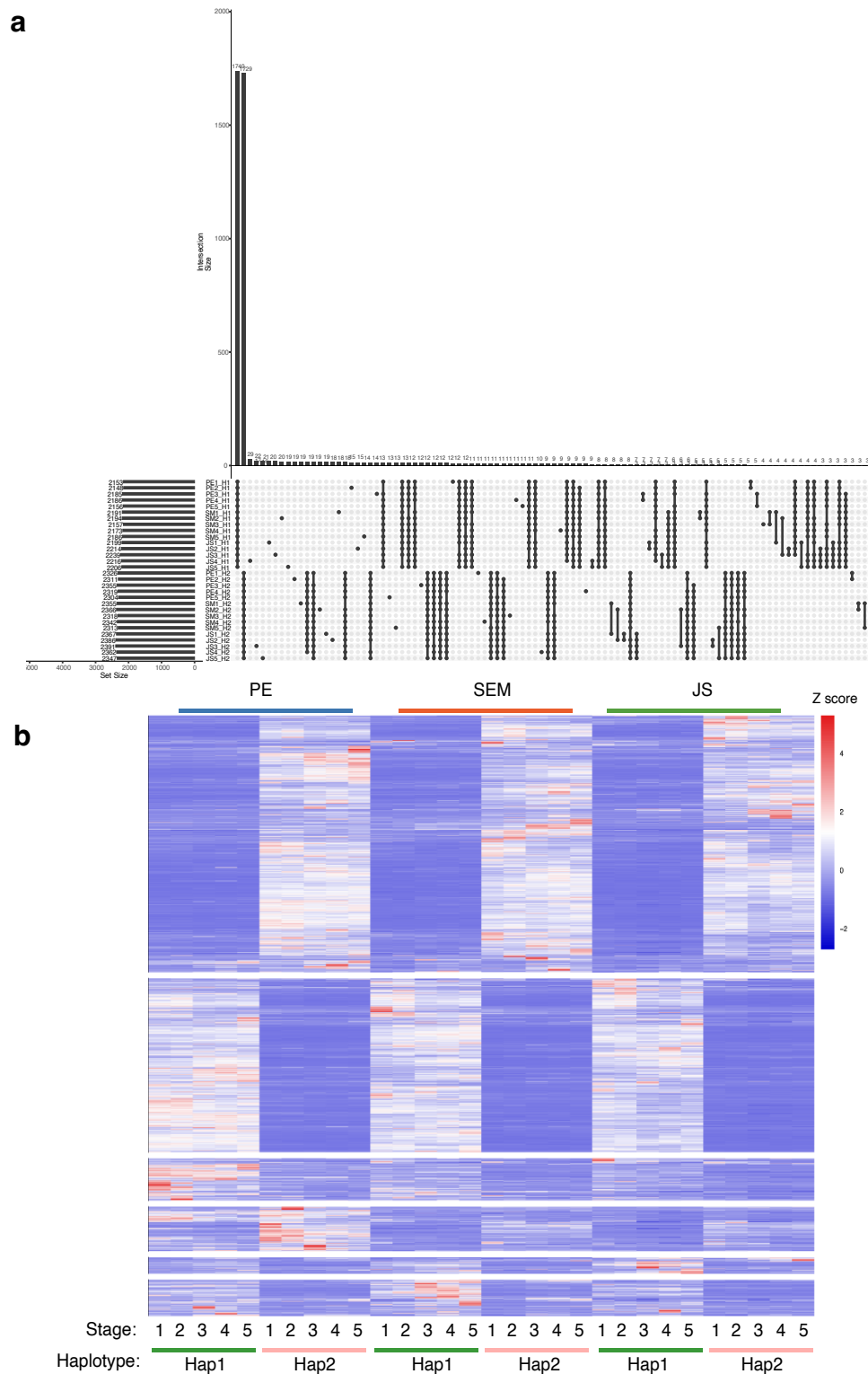


Fig. S13. Allele-specific gene expression during GYH fruit development.

(a) UpSet plot showing overlap of ASEGs in different stages and tissues. Only top 50 categories are shown. H1, the dominant allele was from Hap1. H2, the dominant allele was from Hap2. Sample information is shown in Supplemental Table 7.

(b) The heatmap shows allele-specific expression pattern in 15 samples. The number of dominant alleles of ASEGs.

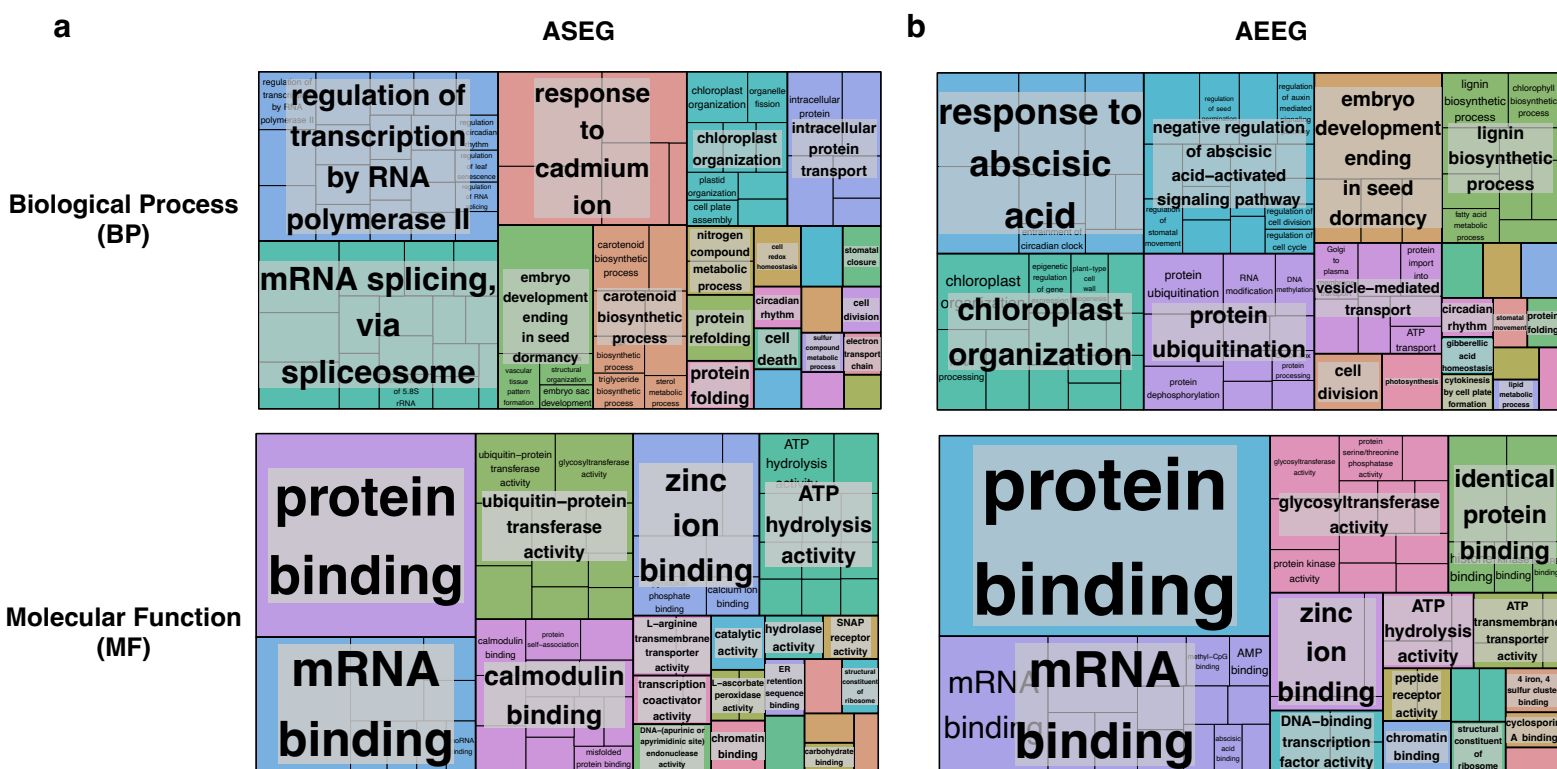


Fig. S14. GO enrichment analysis of ASE and AEE genes The results of the GO term ($P < 0.05$) enrichment analysis visualized by the 'treemap' view of REVIGO. Each rectangle is a single cluster representative. The representatives are joined into 'superclusters' of loosely related terms, visualized with different colours. The size of the rectangles is adjusted to reflect the P-value.