

Telomere-to-telomere genome assemblies and population resequencing of diploid and allotetraploid peanut varieties

In the format provided by the
authors and unedited

Supplementary Figures

Supplementary Fig. 1 Hi-C chromatin interaction map of the six assembly peanut genomes.

Supplementary Fig. 2 Merqury copy number spectrum plots for the six assembled peanut genomes.

Supplementary Fig. 3 Genome synteny comparison between this study and published accessions.

Supplementary Fig. 4 Different insertion events of *Gypsy*-type LTRs in diploid and tetraploid genomes.

Supplementary Fig. 5 Different insertion times of *Gypsy* and *Copia*-type LTRs in different chromosomes.

Supplementary Fig. 6 Structural Variations Among Peanut Genomes.

Supplementary Fig. 7 Transposable elements (TEs) and their association with structural variations (SVs) based on the peanut genome.

Supplementary Fig. 8 PCR validations for translocations, inversions, and duplications identified among different genomes.

Supplementary Fig. 9 Collinearity analysis among peanut assemblies.

Supplementary Fig. 10 Syntenic analyses between Ad vs Bd, At vs Ad and Bt vs Bd.

Supplementary Fig. 11 Structural variation (SV) among different botanical varieties.

Supplementary Fig. 12 The expression pattern of structural variation (SV) genes present on Chromosome 14 in the S83 accession across different tissues.

Supplementary Fig. 13 Phylogenetic tree of 521 peanut accessions.

Supplementary Fig. 14 Population structure and demographic history of peanut populations.

Supplementary Fig. 15 LTR insertion on selected sweep regions.

Supplementary Fig. 16 Selected sweep and introgression among different populations.

Supplementary Fig. 17 SNP Variation Types in Selected and Introgressed Regions.

Supplementary Fig. 18 Percentage of share and specific variation between diploids and tetraploids.

Supplementary Fig. 19 Enrichment analyses of significantly selected and introgression among different populations.

Supplementary Fig. 20 Tissue-specific expression of *AhWRI1* gene.

Supplementary Fig. 21 The expression levels of *AhFUS3*, *AhABI3*, *AhLECs*, and their target gene *AhWRI1* in high and low oil content peanut accessions.

Supplementary Fig. 22 The oil content identification of overexpressing *AhWRI1* in *Glycine max* plant.

Supplementary Fig. 23 Tissue-specific expression of *AhGSA1* gene.

Supplementary Fig. 24 WGCNA analysis and network construction.

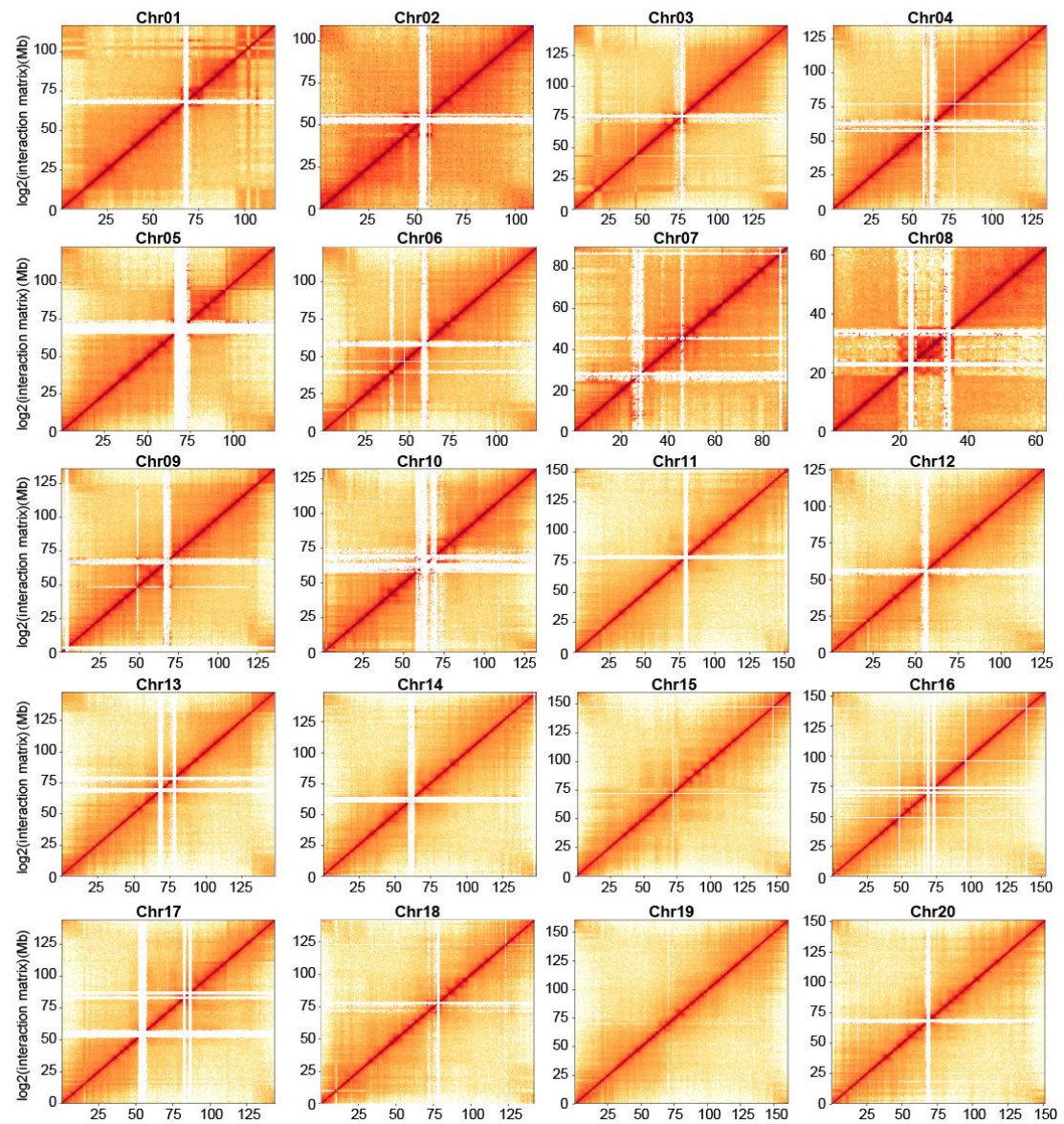
Supplementary Fig. 25 Evolutionary tree of transcription factors and enzymes in the peanut genome.

Supplementary Fig. 26 The flowchart of genome assembly.

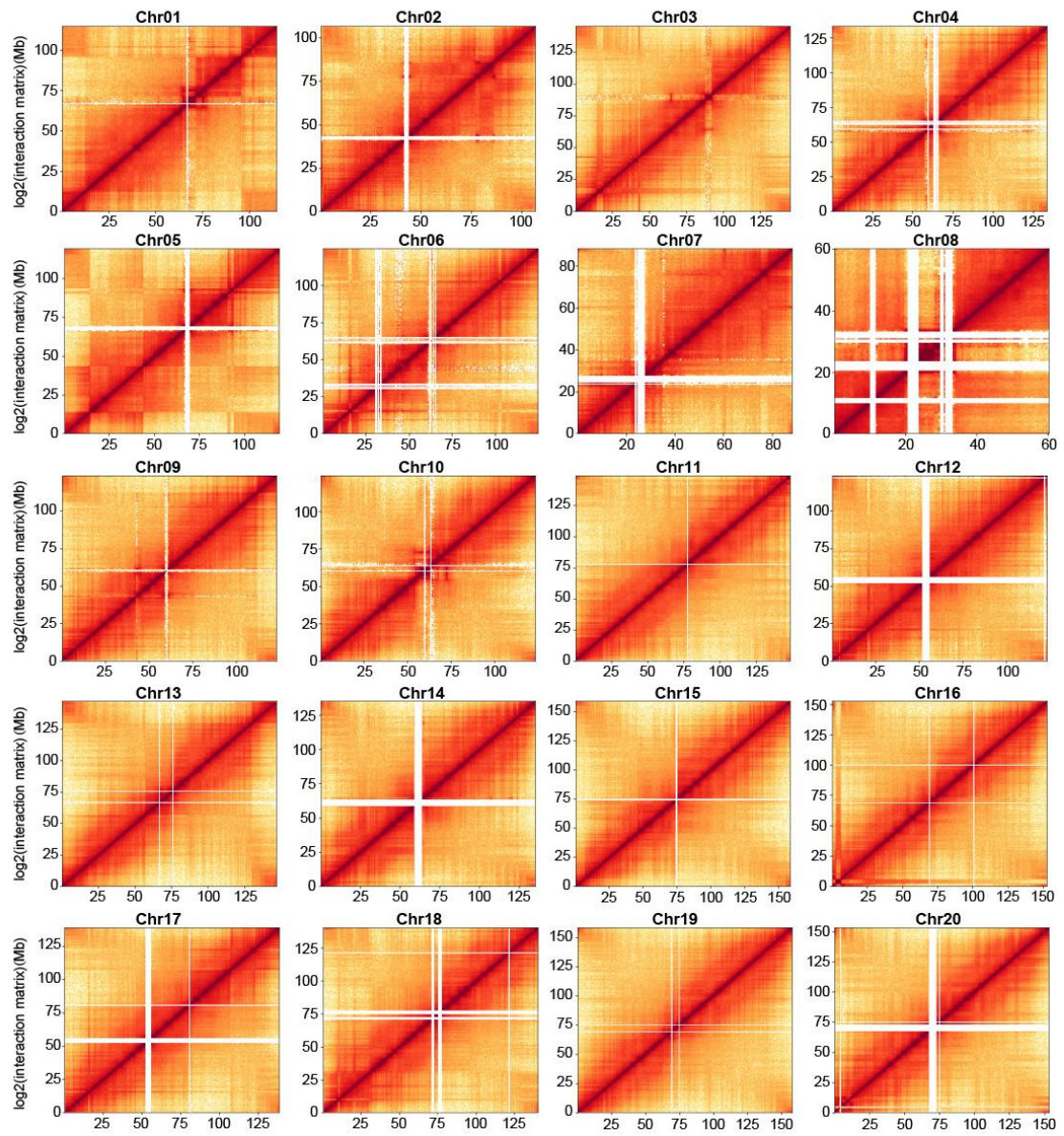
Supplementary Fig. 27 GWAS for 33 traits.

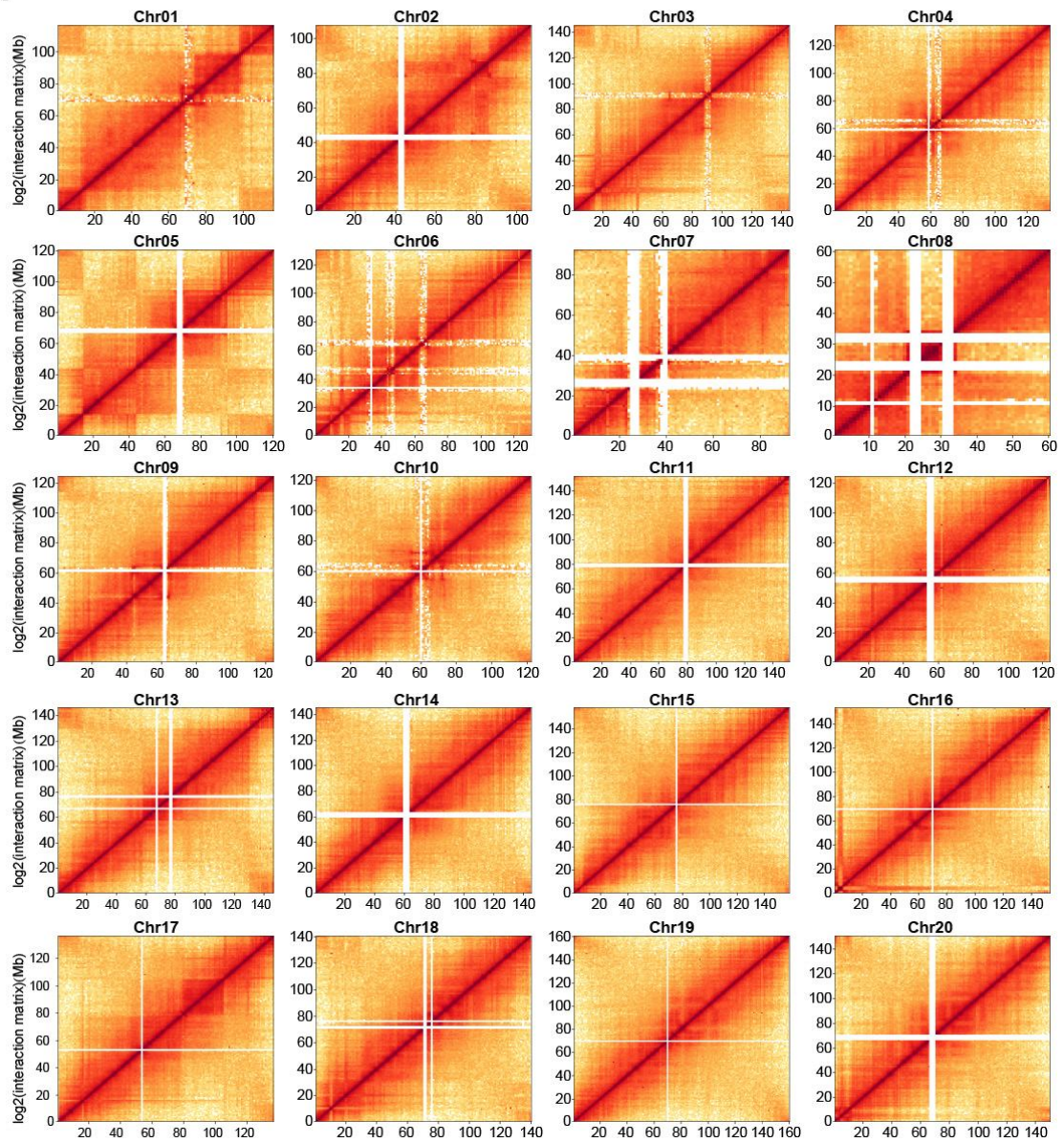
Supplementary Notes

a

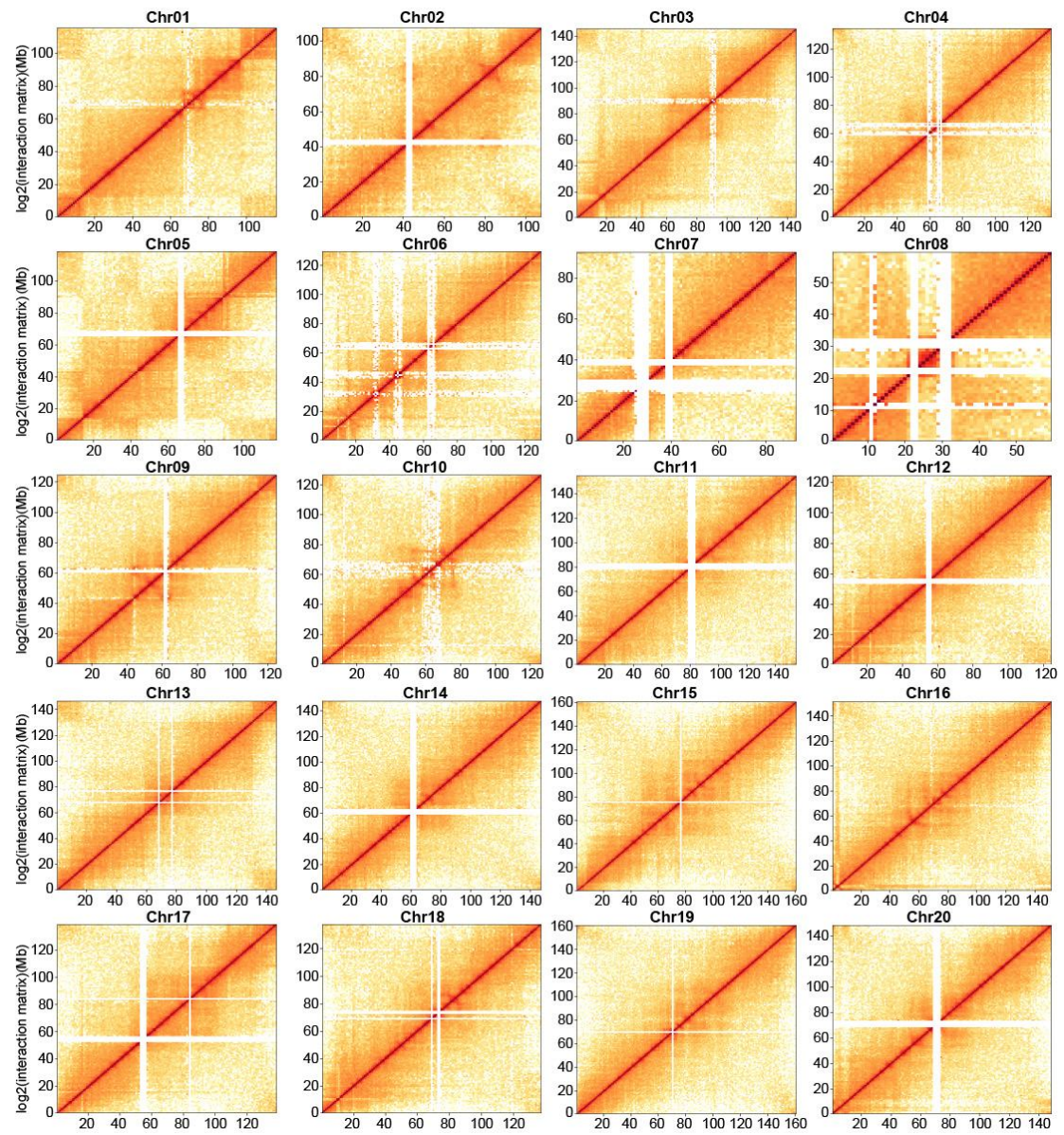


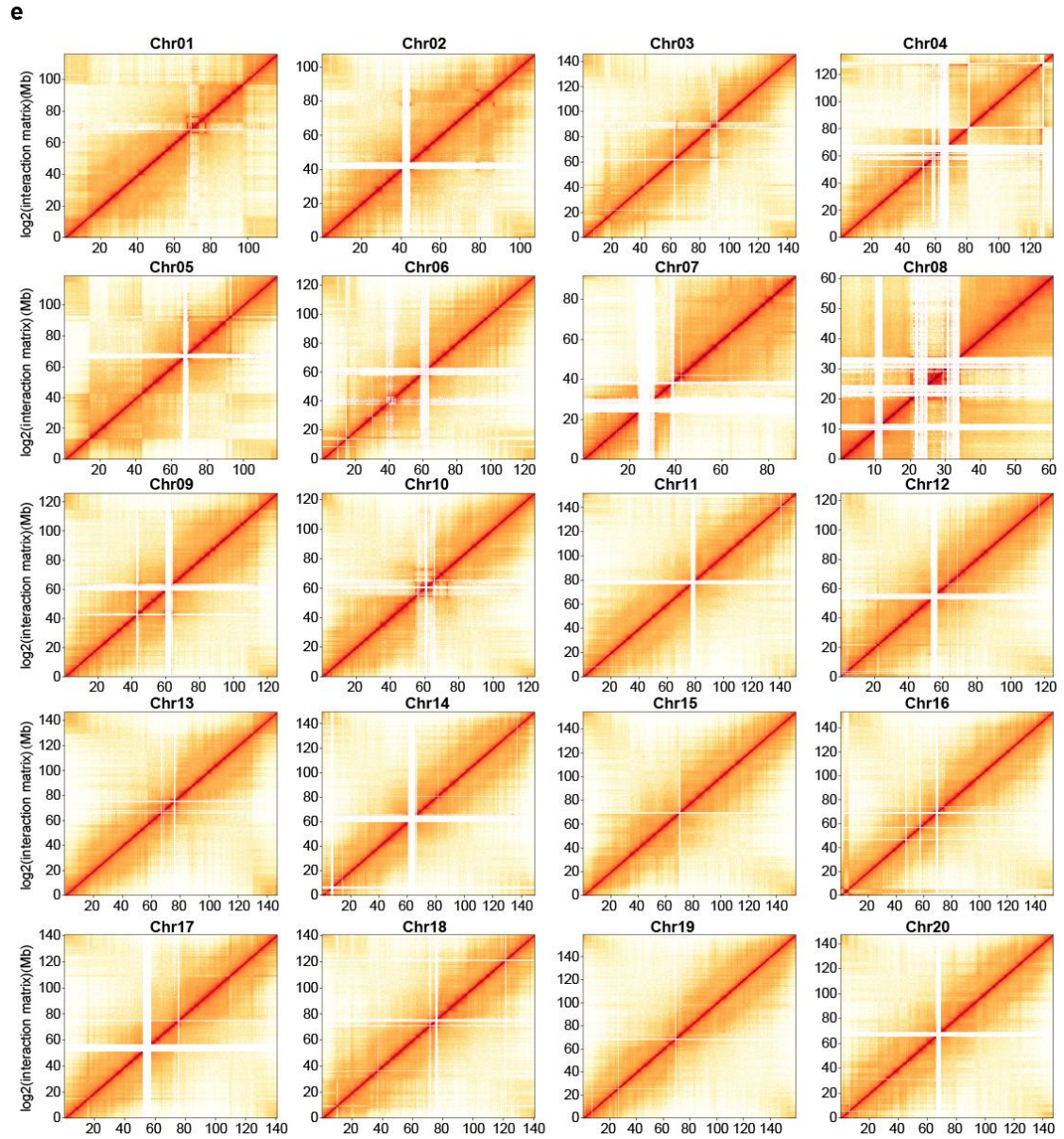
b



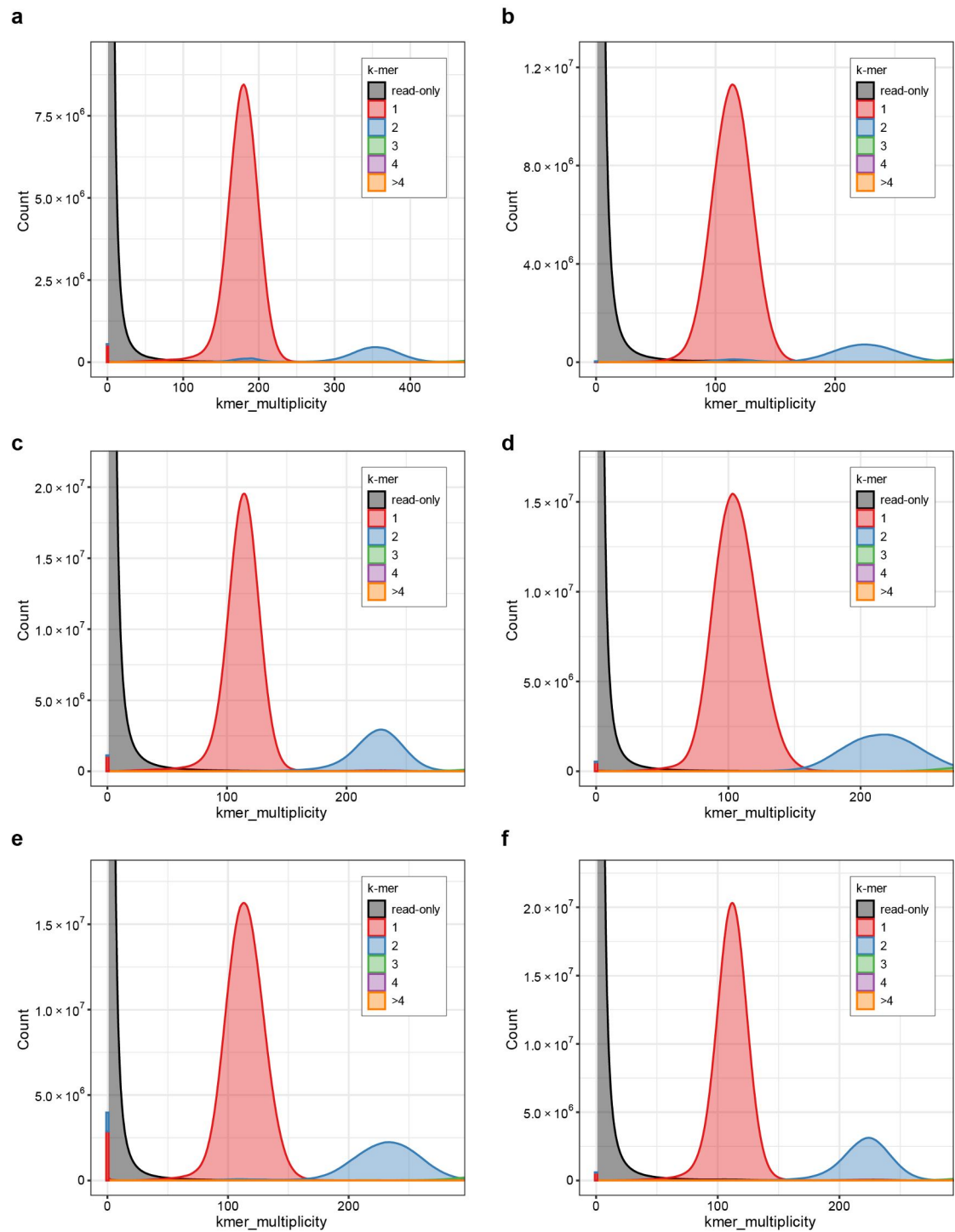
C

d

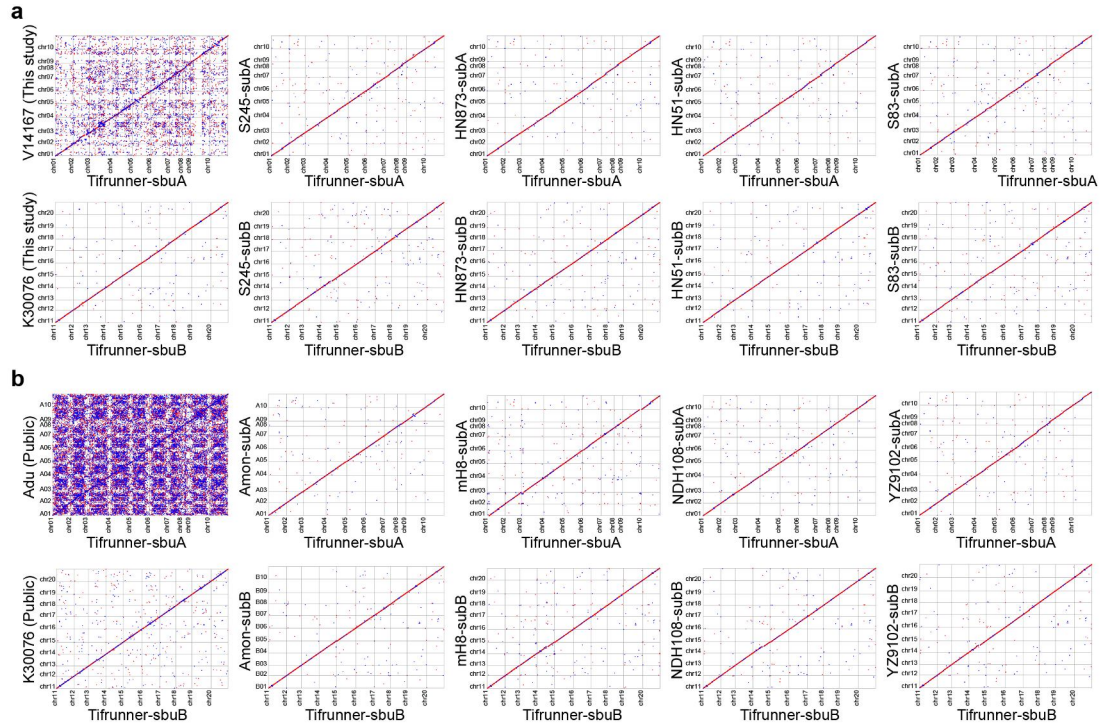




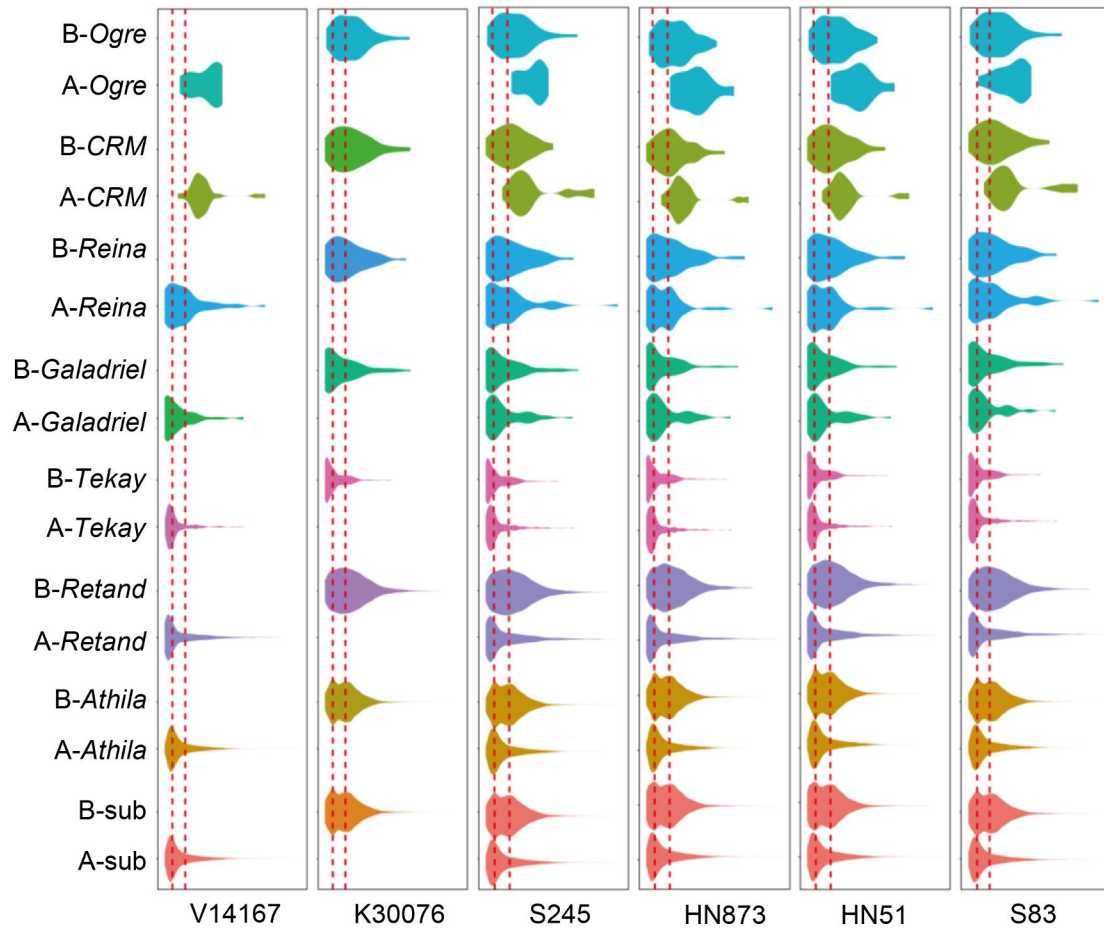
Supplementary Fig. 1 Hi-C chromatin interaction map of the six assembly peanut genomes. Synthetic intra-chromosomal matrix of chr01 to chr20, respectively. **a**, V14167(Chr01-Chr10) and K30076(Chr11-Chr20) accessions. **b-e**, S245, HN873, HN51 and S83 accessions. The colored bar on the right side of the Figure indicates the strength of interaction.



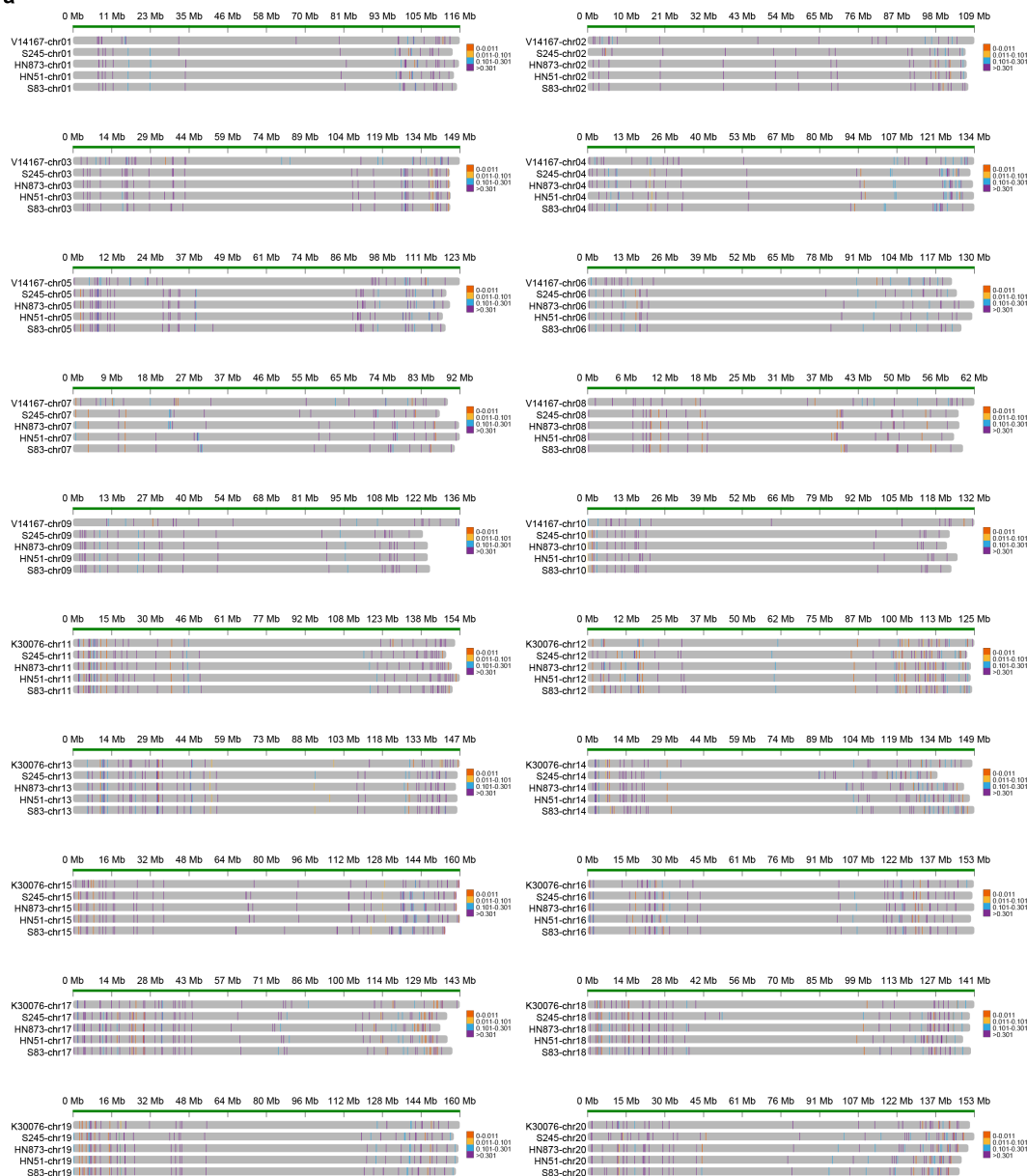
Supplementary Fig. 2 Merqury copy number spectrum plots for the six assembled peanut genomes. The copy number spectrum (spectra-cn) of the same K-mers is plotted as stacked histograms, colored by the copy numbers found in the combined assembly. The samples are: **a**, V14167; **b**, K30076; **c**, S245; **d**, HN873; **e**, HN51; **f**, S83.



Supplementary Fig. 3 Genome synteny comparison between this study and published accessions. Genomic synteny analysis between the A genome/subgenome vs. Tifrunner_subA and B genome/subgenome vs. Tifrunner_subB, with Tifrunner as the reference (Bertioli et al. (2019)¹). **a**, All six T2T genome assemblies in this study. **b**, Six genomes from previously published articles, including the K30076 genome from Bertioli et al. (2016)². The Adu, Amon, mH8, and NHD108 genomes are from Zhao et al. (2025)³; YZ9102 is from Wang et al. (2025)⁴. Except for YZ9102, which is a T2T genome, the contig sequences in the other five genomes were deleted prior to alignment, leaving only chromosome sequences.



Supplementary Fig. 4 Different insertion events of Gypsy-type LTRs in diploid and tetraploid genomes. Gypsy-type LTRs divided into seven subfamilies, including: *Athila*, *Retand*, *Tekay*, *Galadriel*, *Reina*, *CRM* and *Ogre*. The peak represents an expansion event, with the dashed line on the right indicating the first expansion occurring around 0.69 million years ago (Mya), and the dashed line on the left indicating the second expansion occurring around 0.27 Mya.

a

b

C



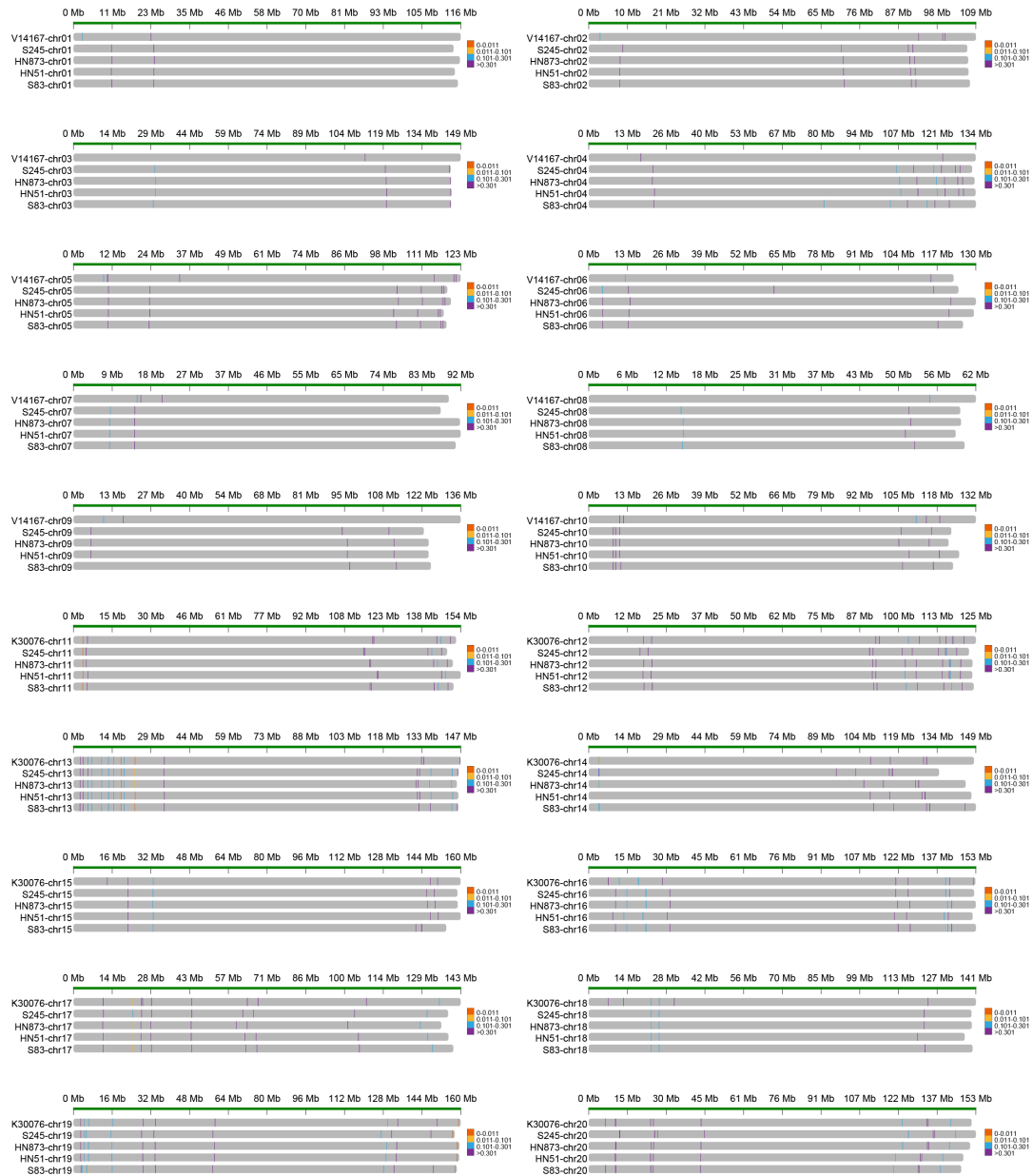
d



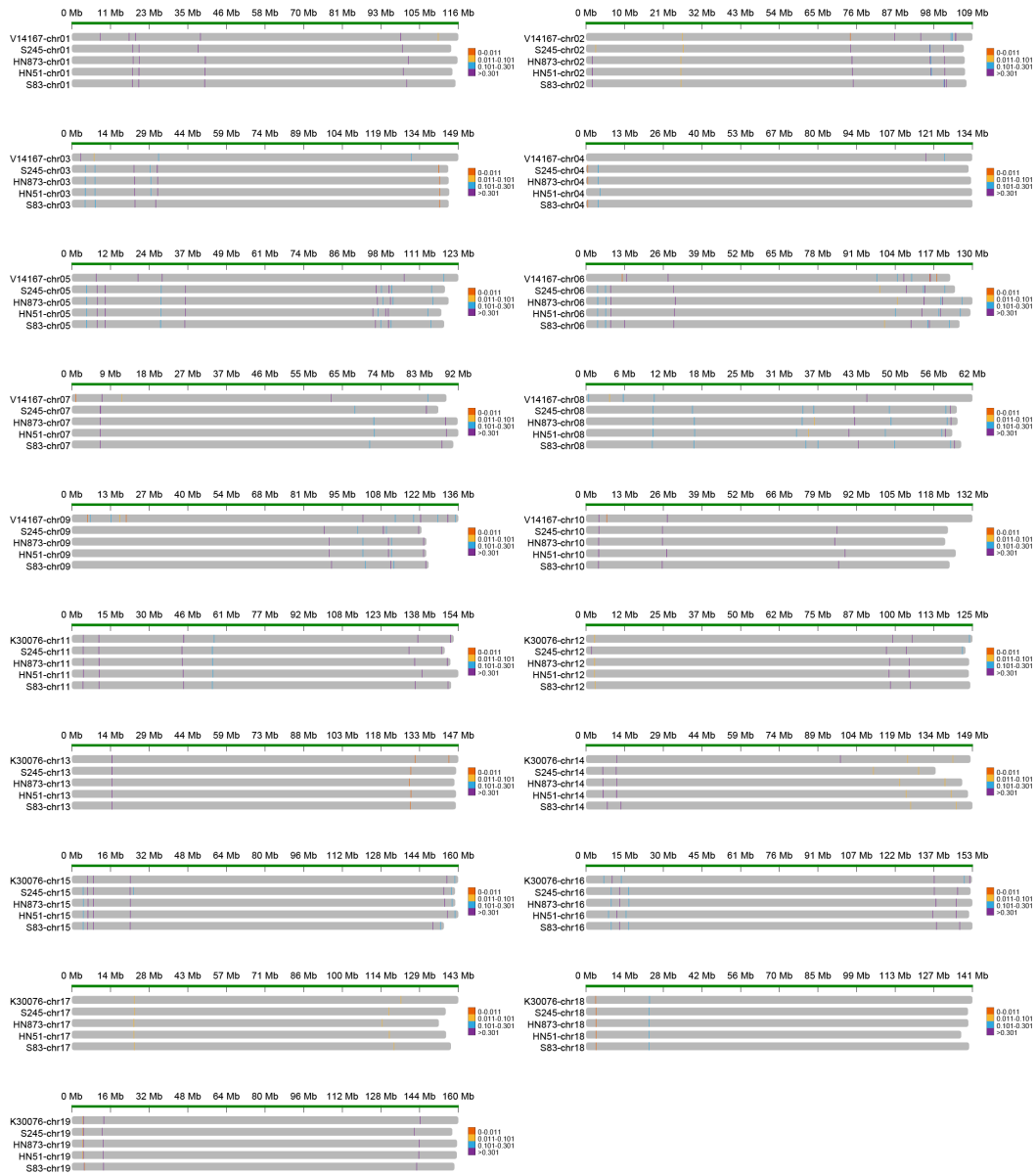
e



f



g



h

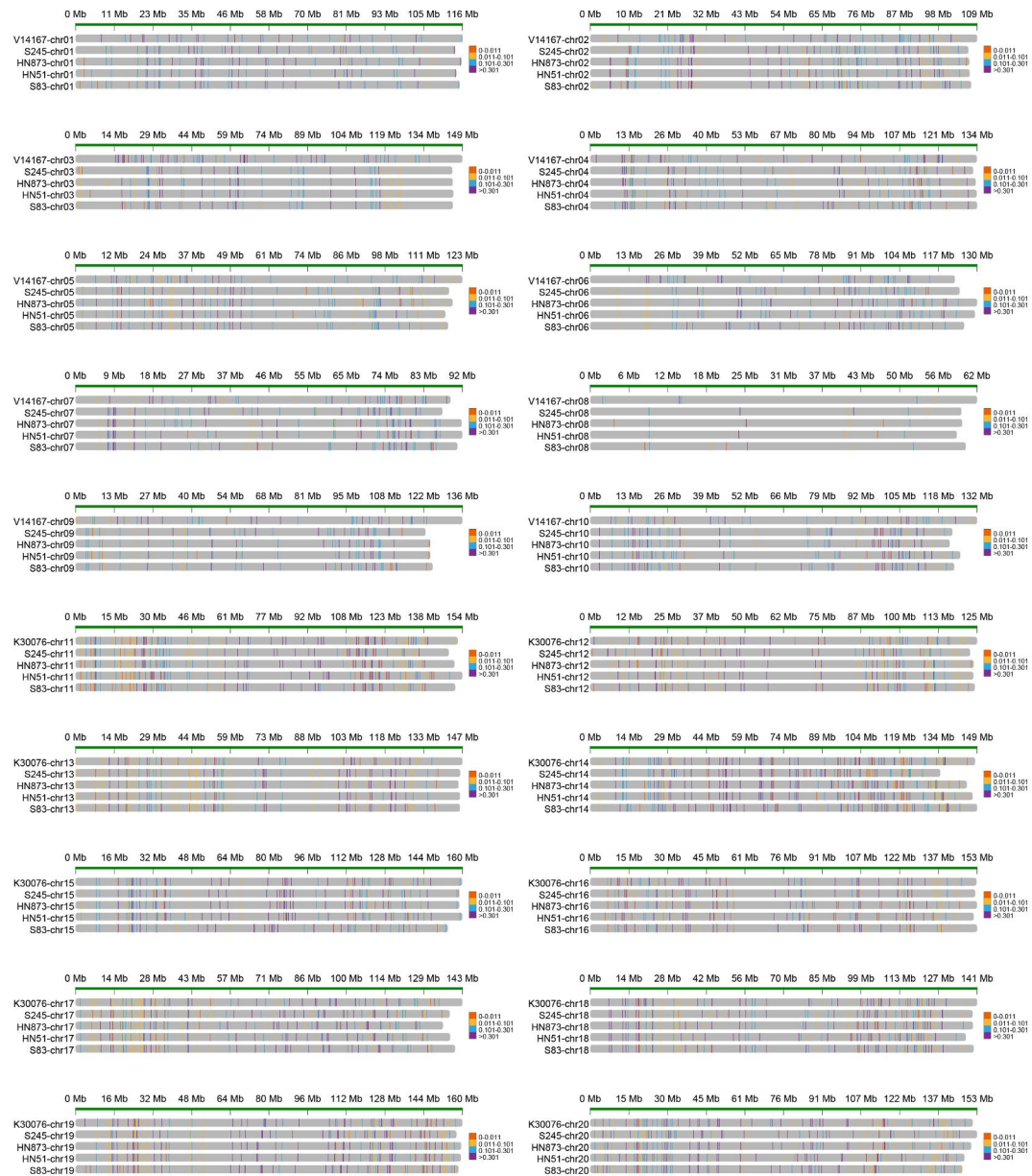


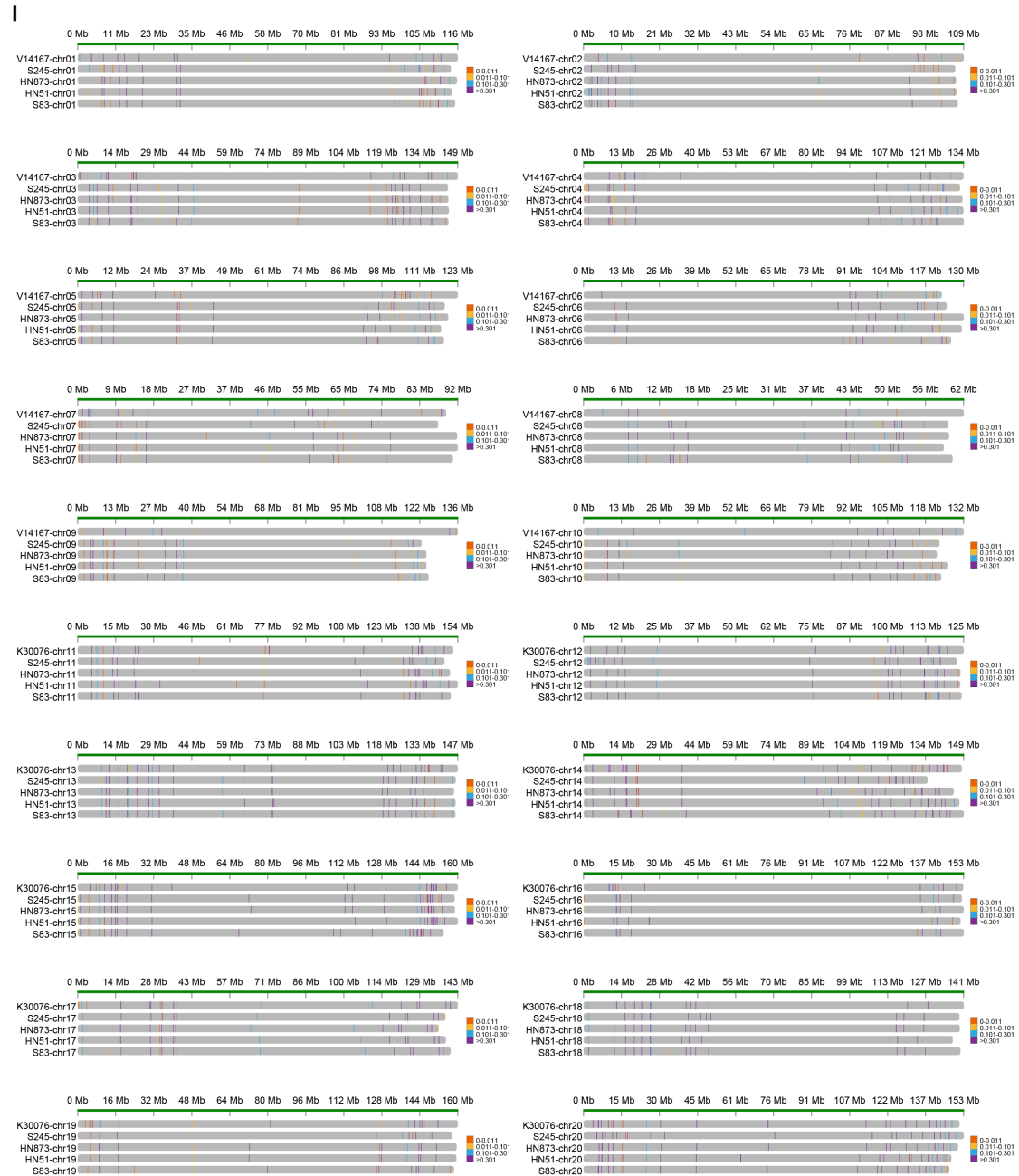


j

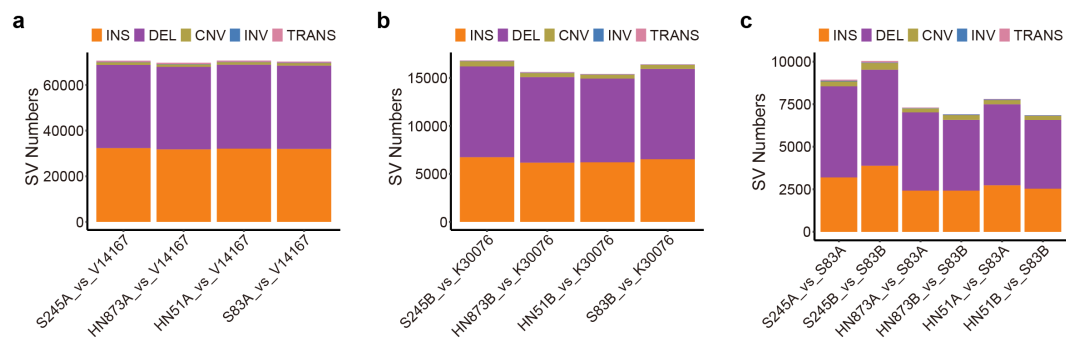


k

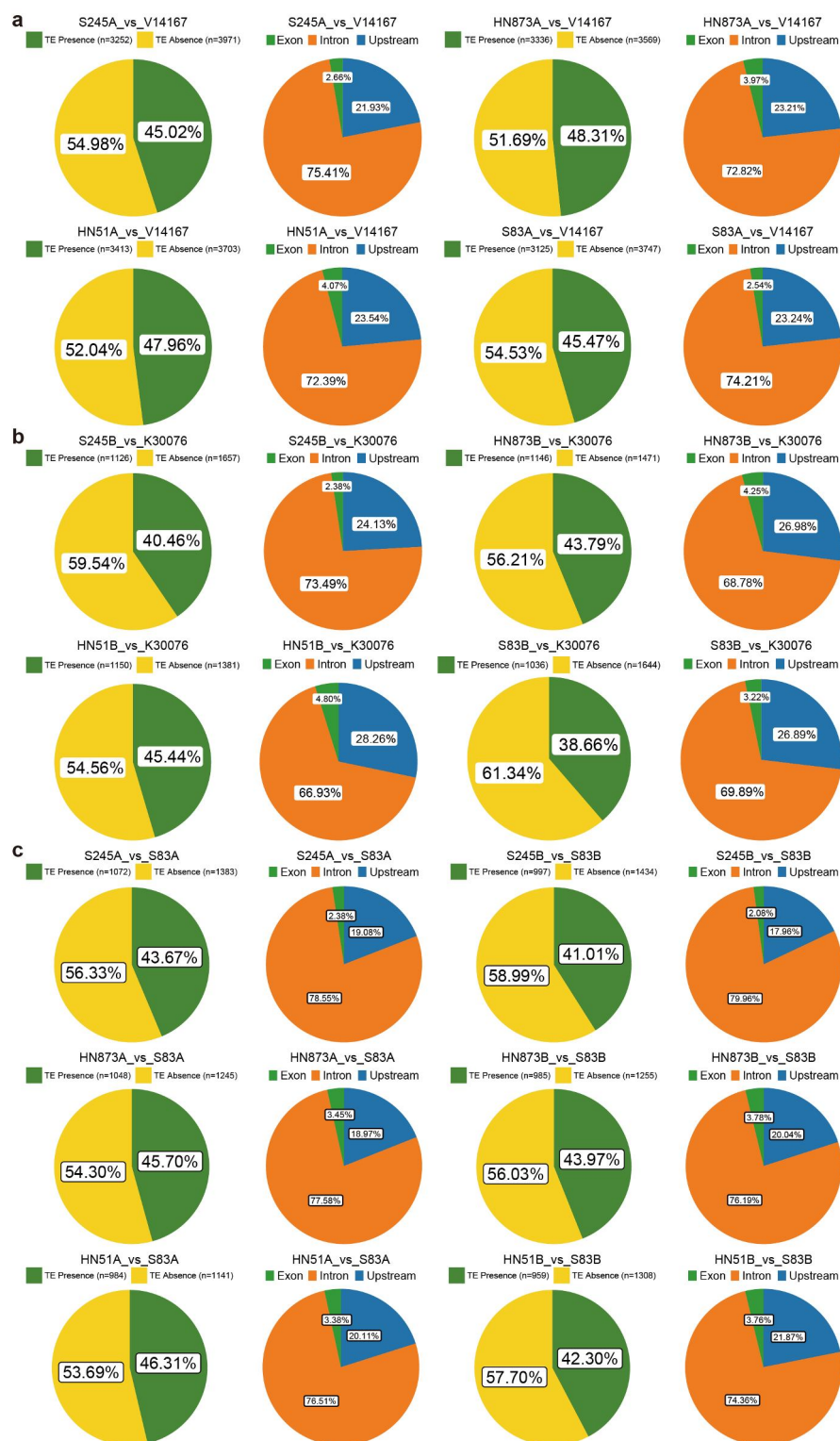




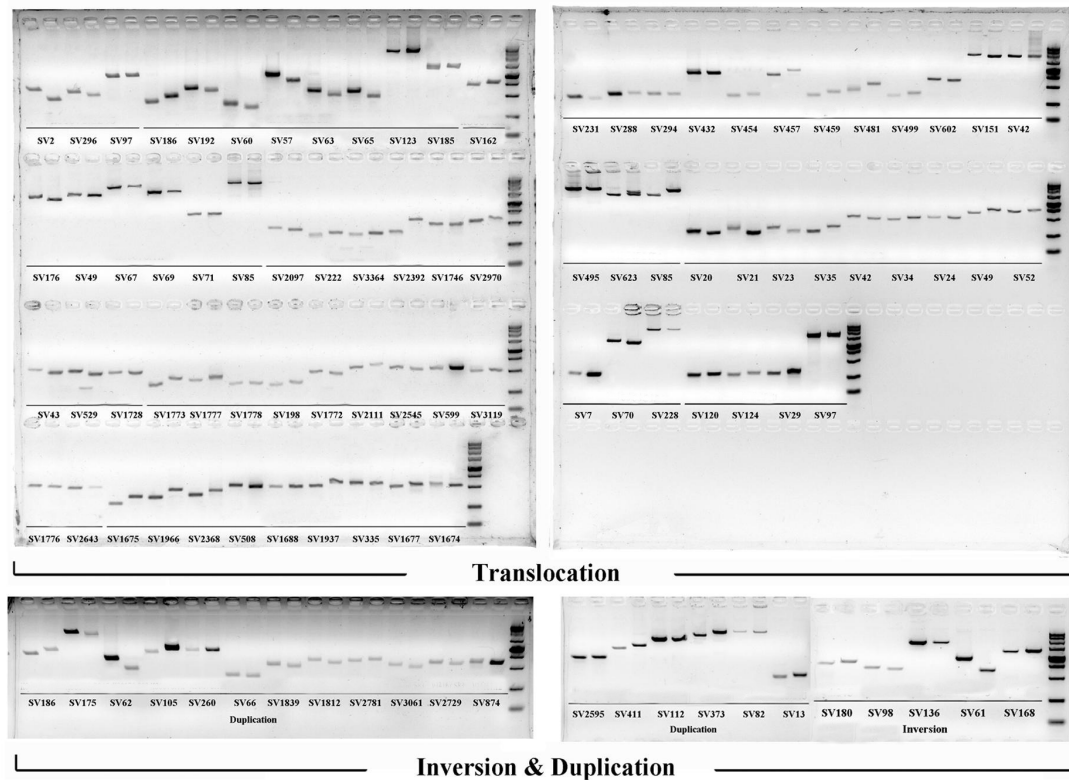
Supplementary Fig. 5 Different insertion times of Gypsy and Copia-type LTRs in different chromosomes. LTRs were divided into 12 subfamilies, including: **a**, *Ale*; **b**, *Athila*, **c**, *Bianca*; **d**, *CRM*; **e**, *Galadriel*; **f**, *Ikeros*; **g**, *Ivana*; **h**, *Reina*; **i**, *Retand*; **j**, *SIRE*; **k**, *Tekay*; **l**, *Tork*. The insertion times (Mya) are represented by different color.



Supplementary Fig. 6 Structural Variations Among Peanut Genomes. a-c, The number of different types of structural variants present in each genome compared to K30076 (a), V14167 (b), and S83 (c), respectively. The types of structural variants include: INS (Insertion), DEL (Deletion), CNV (Copy Number Variation), INV (Inversion), and TRANS (Translocation).

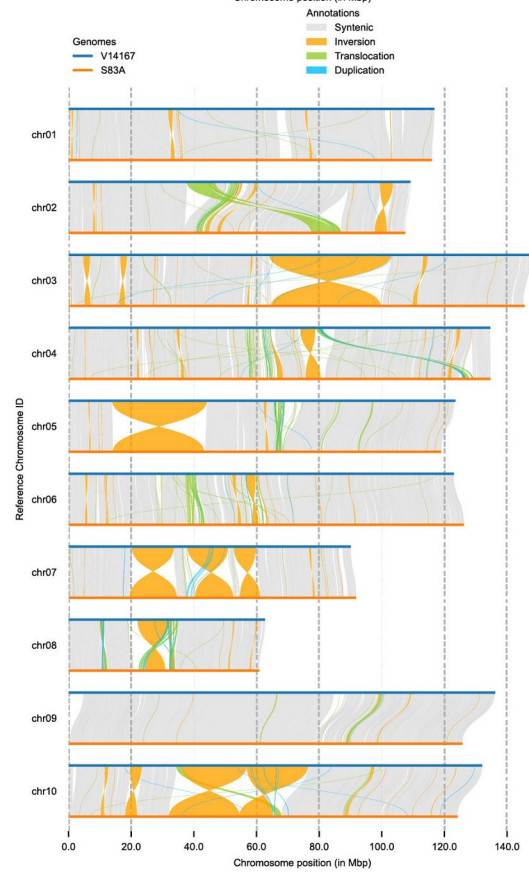
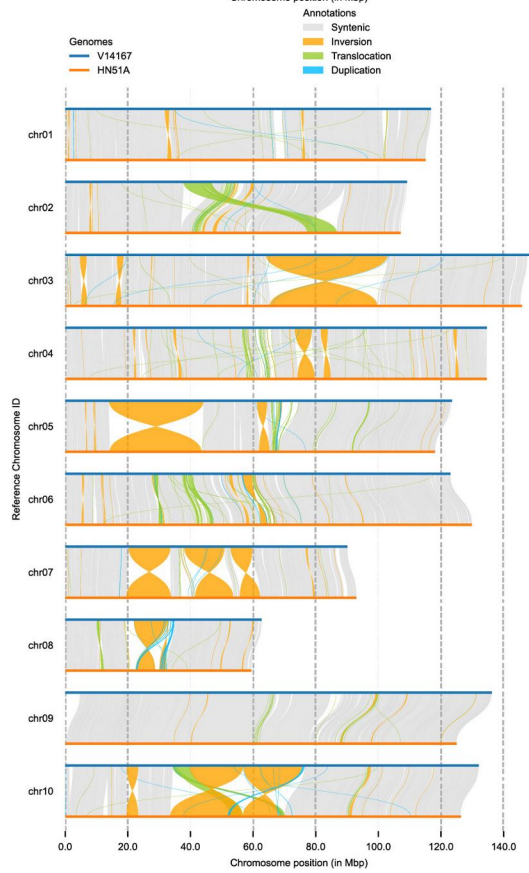
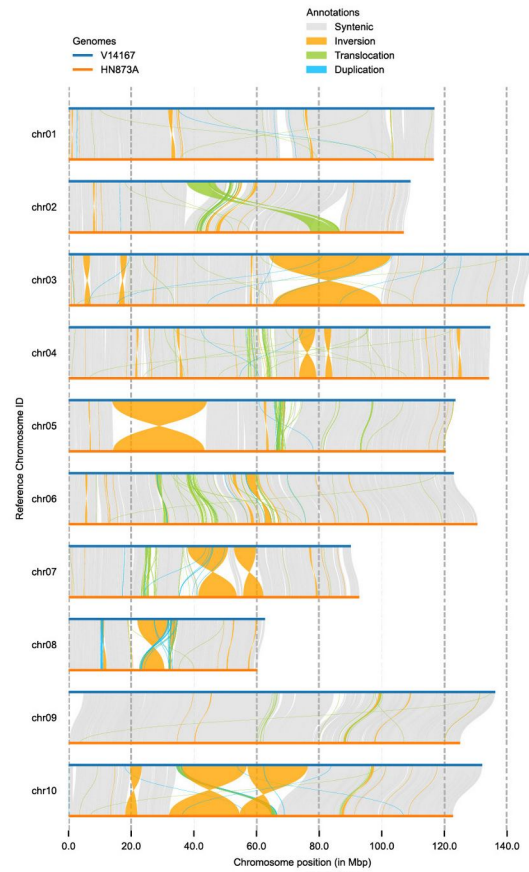
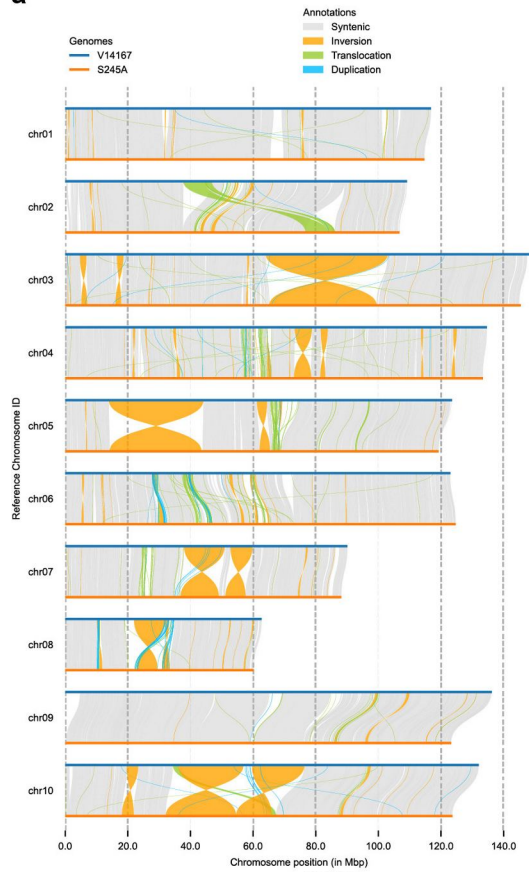


Supplementary Fig. 7 Transposable elements (TEs) and their association with structural variations (SVs) based on the peanut genome. The proportions of SV-genes matched to TEs (on the left), and the percentage of TEs in TE-SV-genes located within 1kb upstream of genes, introns, and exons. **a**, At vs Ad (V14167); **b**, Bt vs Bd (K30076); **c**, At vs At (S83) and Bt vs Bt (S83).

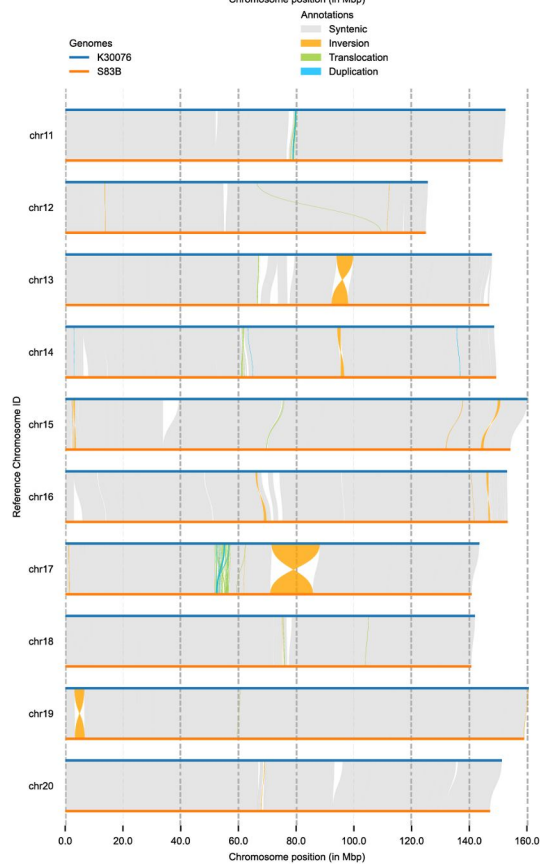
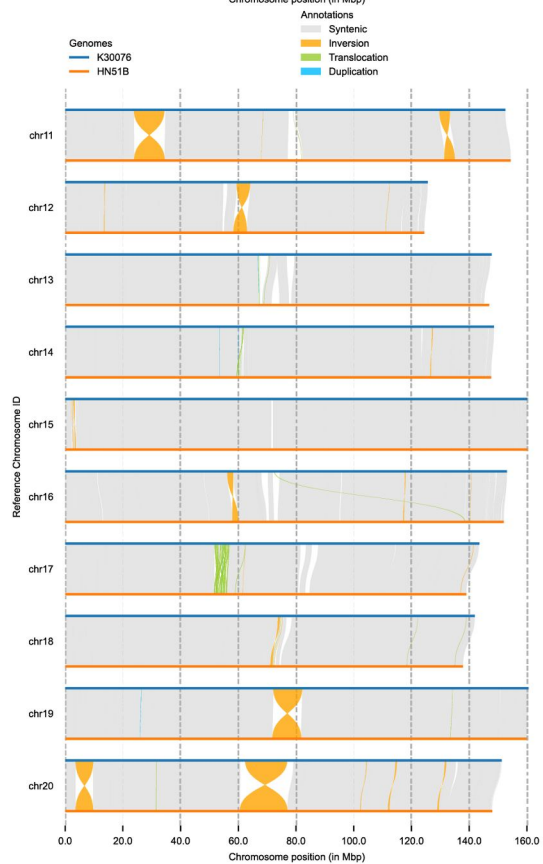
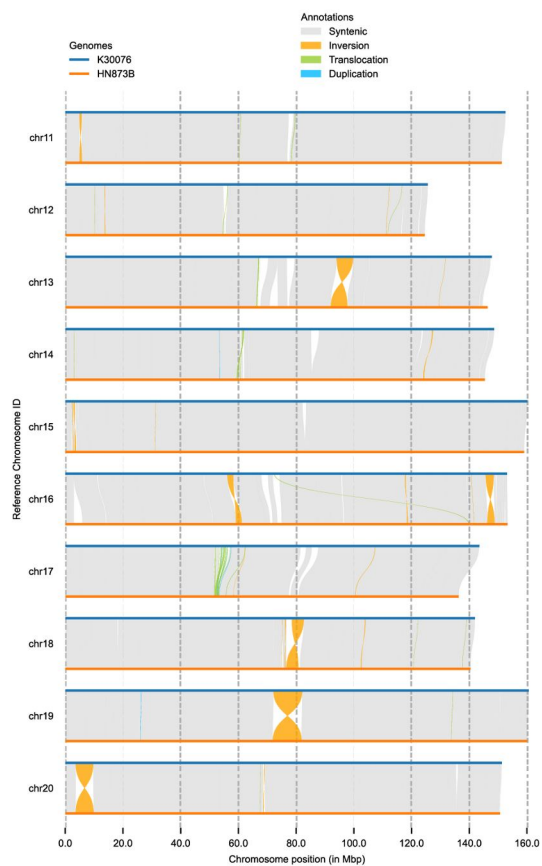
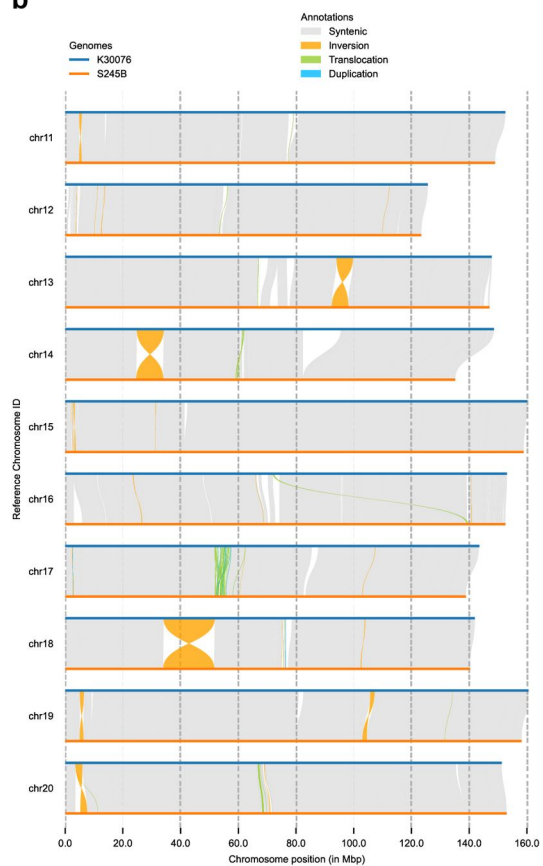


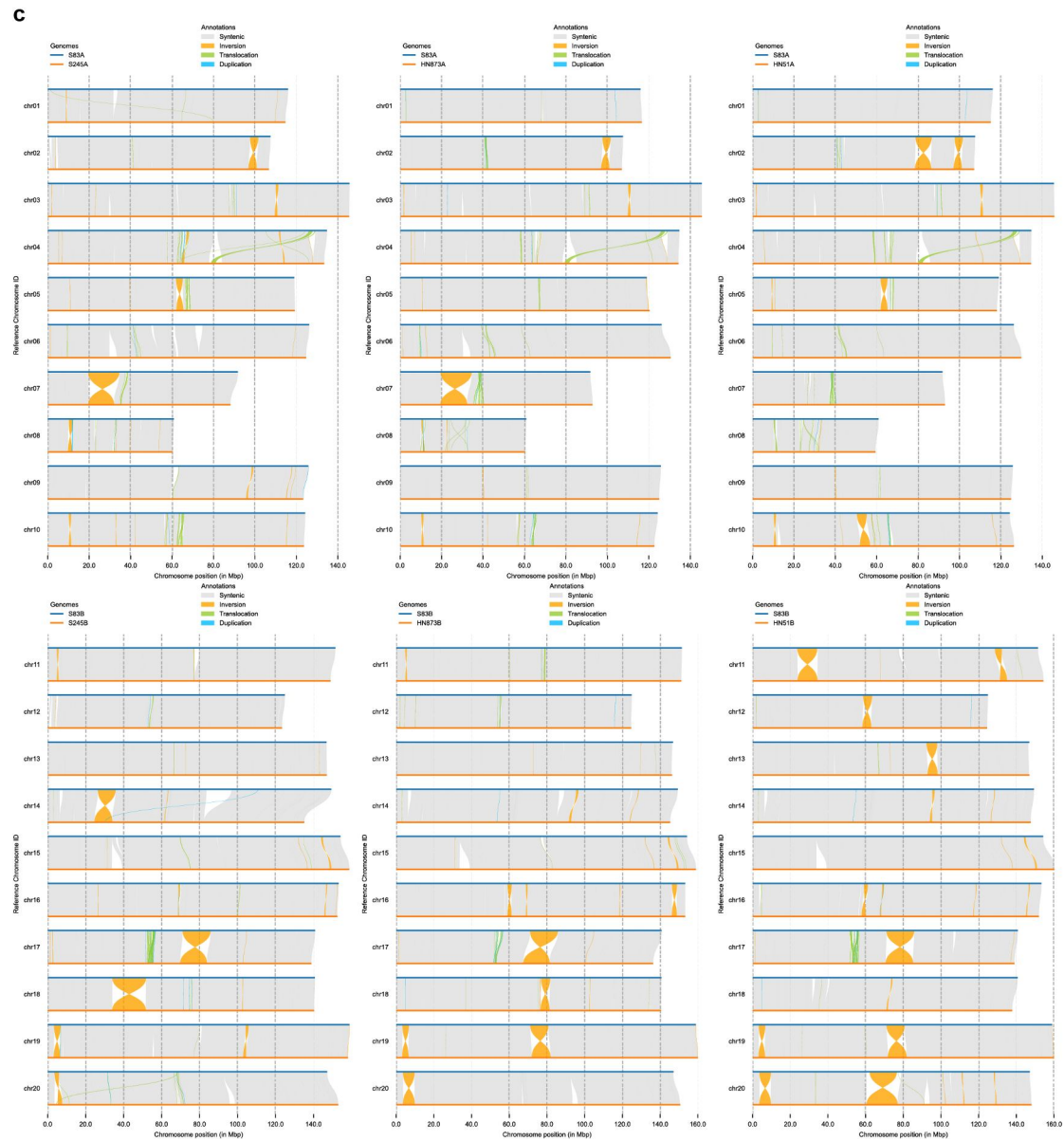
Supplementary Fig. 8 PCR validations for translocations, inversions, and duplications identified among different genomes. Amplified fragments were obtained using primers designed based on the sequence of SV sites within 2 kb upstream and downstream.

a

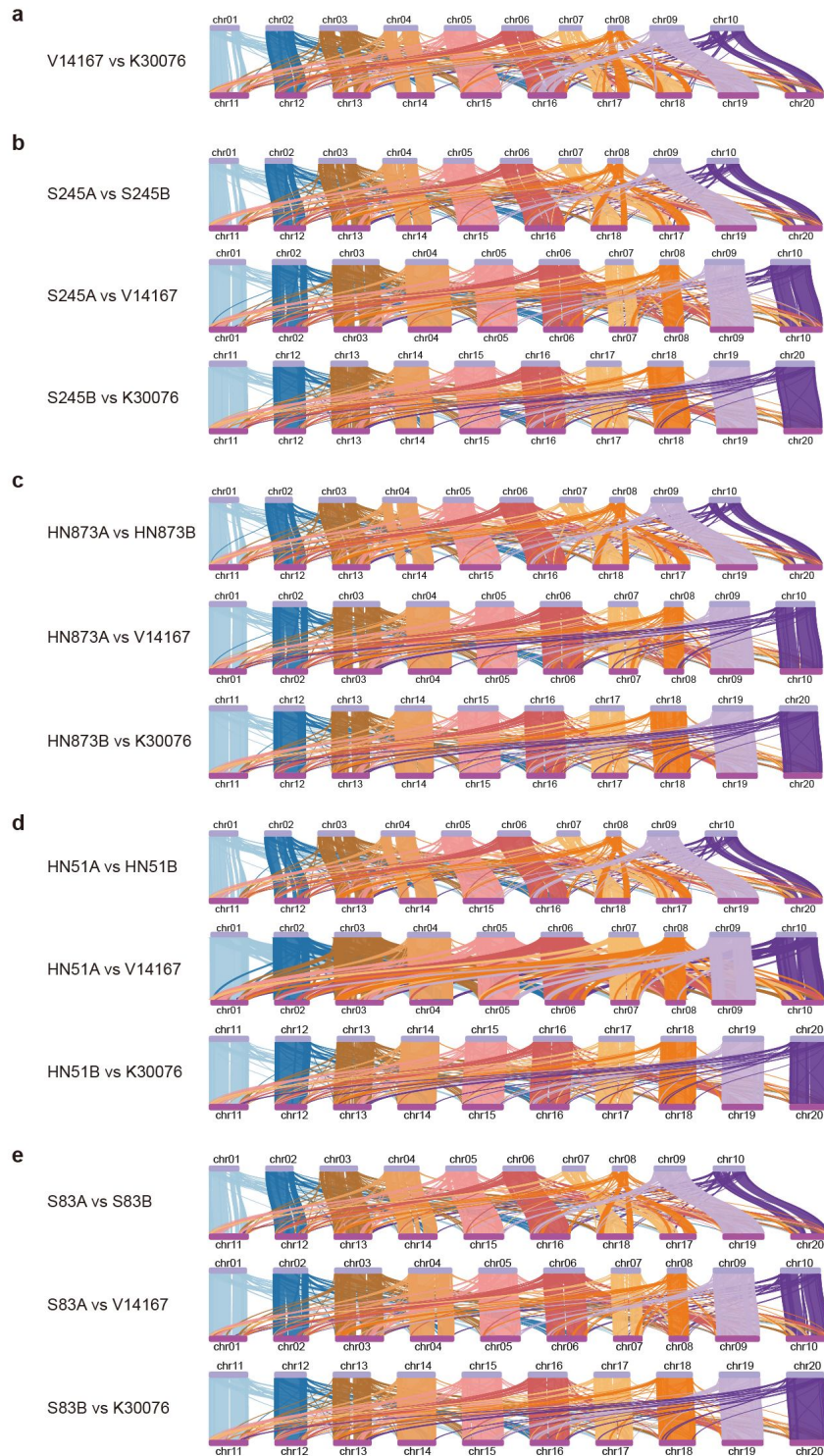


b





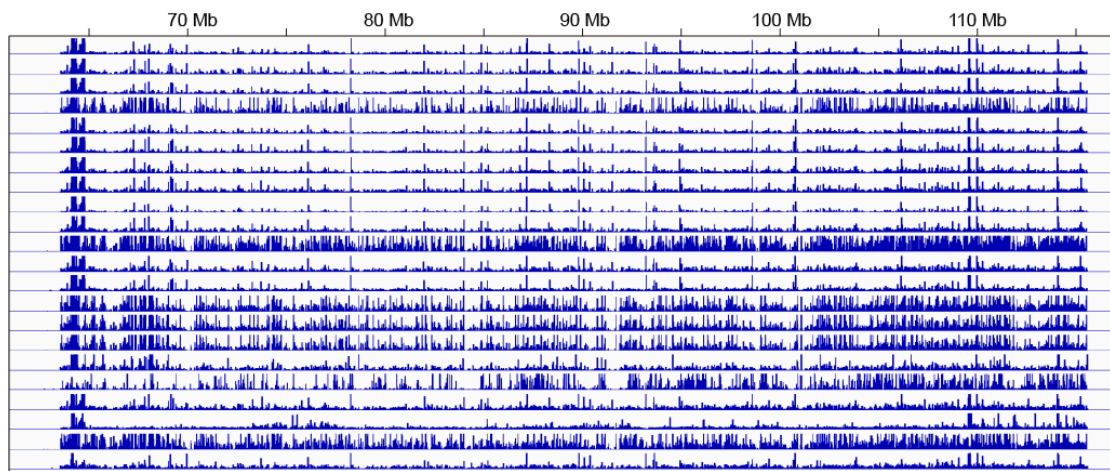
Supplementary Fig. 9 Collinearity analysis among peanut assemblies. a, Genome collinearity of V14167 (AA) genome and At subgenomes, including S245A-vs- V14167, HN873A-vs-V14167, HN51A-vs-V14167 and S83A-vs-V14167; V14167 as reference. **b**, Genome collinearity of K30076 (BB) genome and Bt subgenomes; S245B-vs-K30076, HN873B-vs-K30076, HN51B-vs-K30076 and S83B-vs-K30076; K30076 as reference. **c**, Genome collinearity among tetraploid peanuts, including: S245A-vs-S83A, HN51A-vs-S83A, HN873A-vs-S83A and S245B-vs-S83B, HN51B-vs-S83B, HN873B-vs-S83B, S83A and S83B as reference.

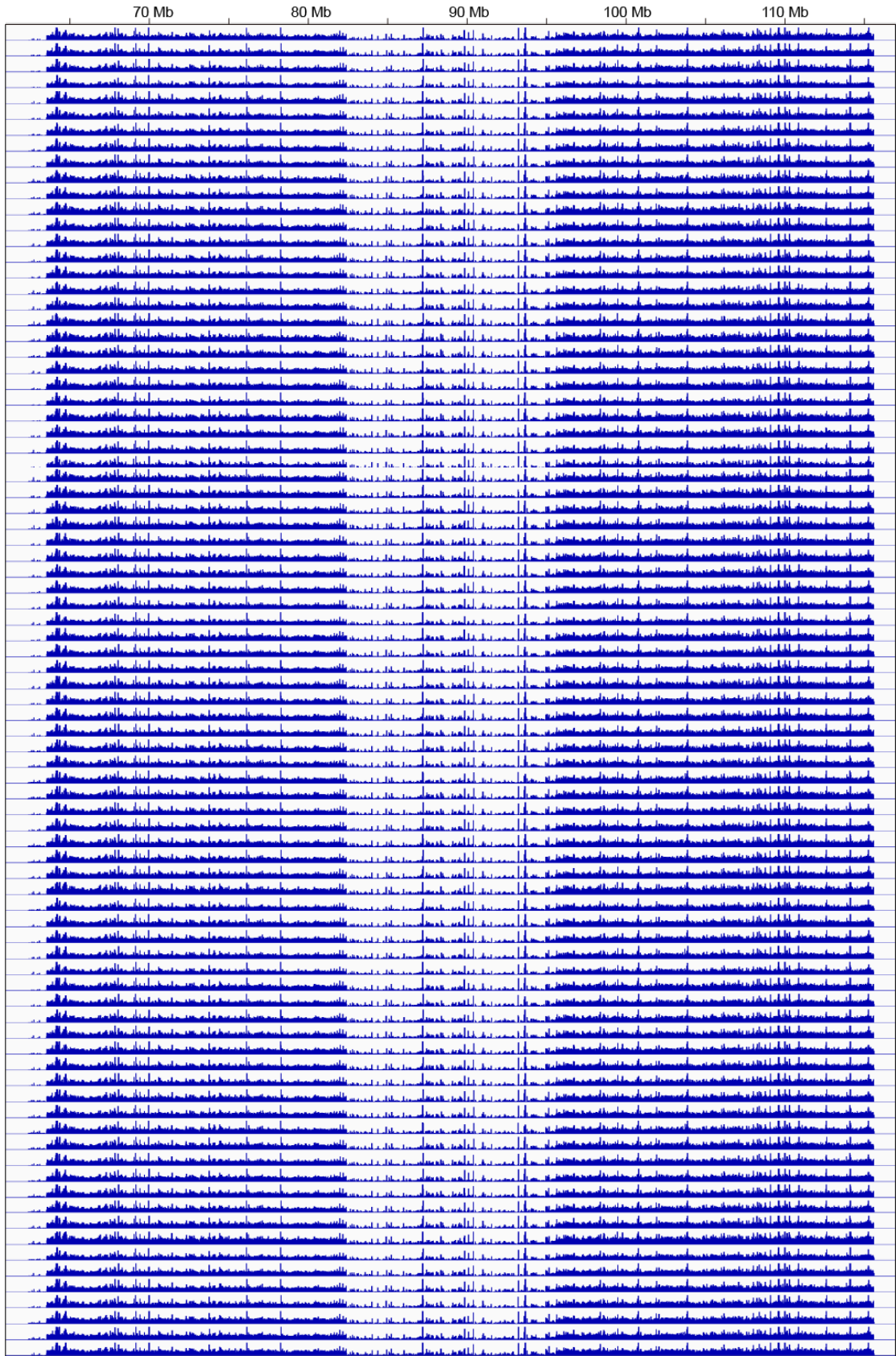


Supplementary Fig. 10 Syntenic analyses between Ad vs Bd, At vs Ad and Bt vs Bd.

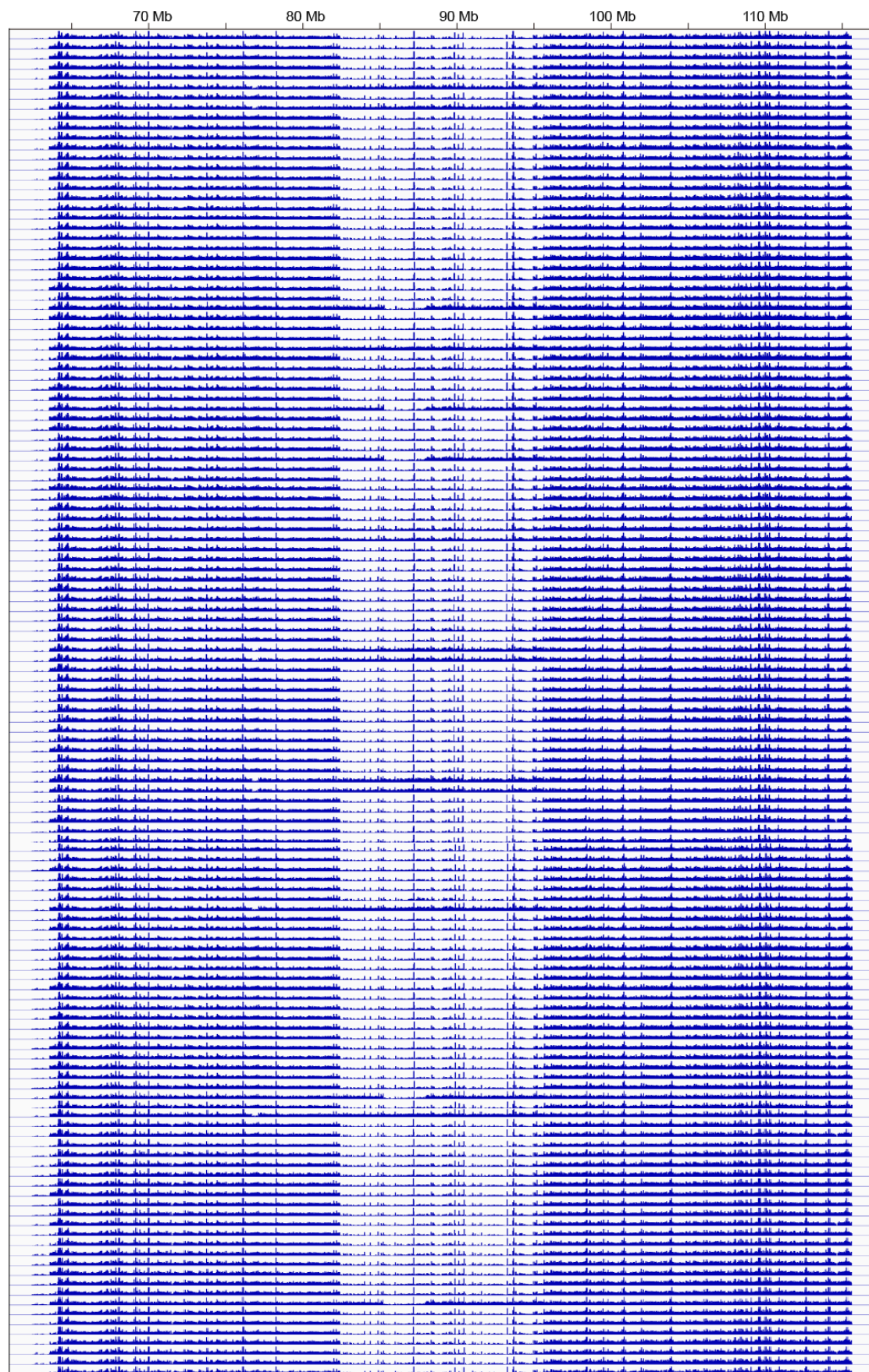
a, Comparison between diploid genomes Ad and Bd. **b-e**, Synteny analysis of the At and Bt subgenomes of *A.hypogaea* and their corresponding ancestral diploids. Homeologous blocks of ≥ 5 gene pairs between Chr01–Chr10 and Chr11–Chr20 are connected with lines.

a

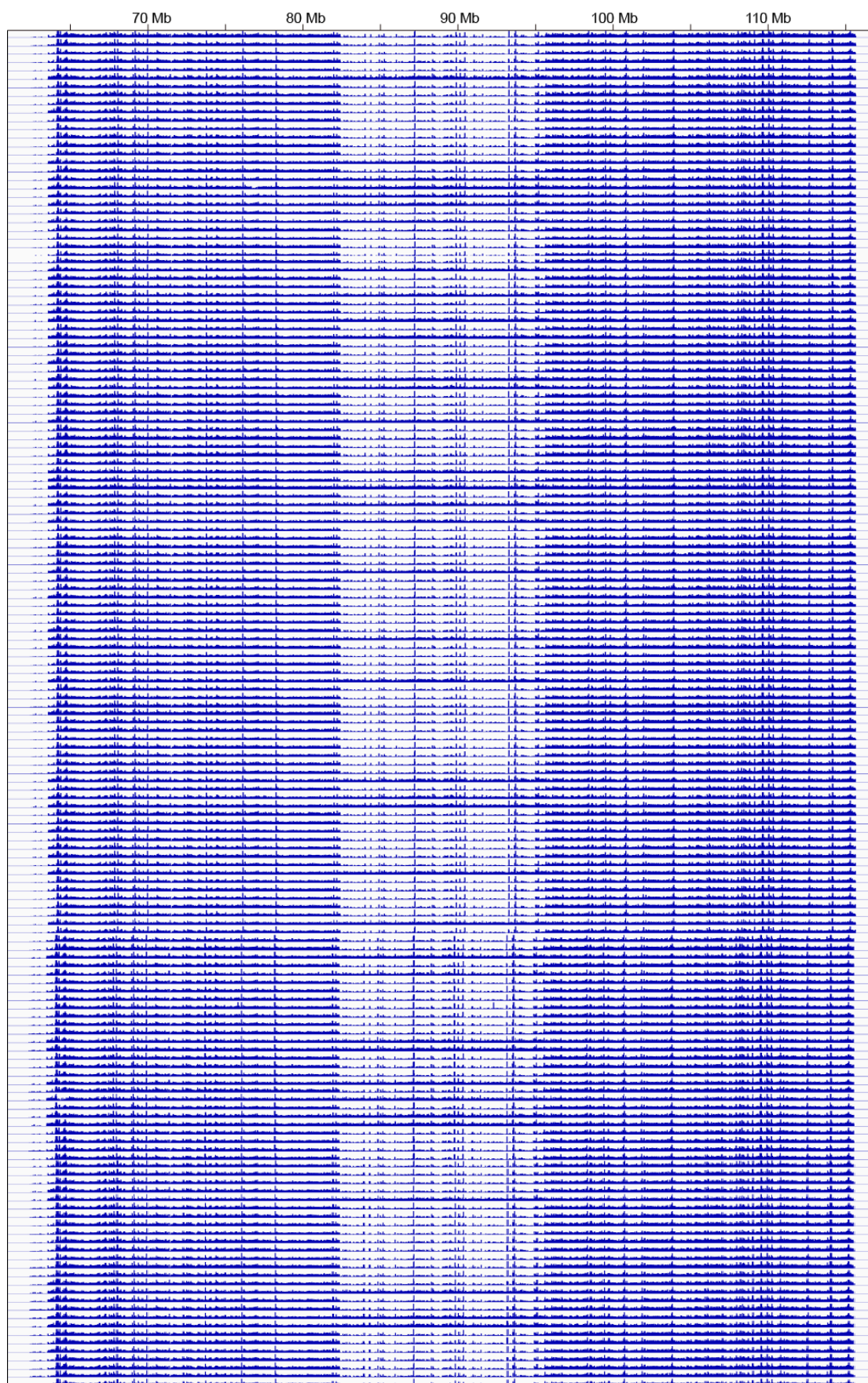


b

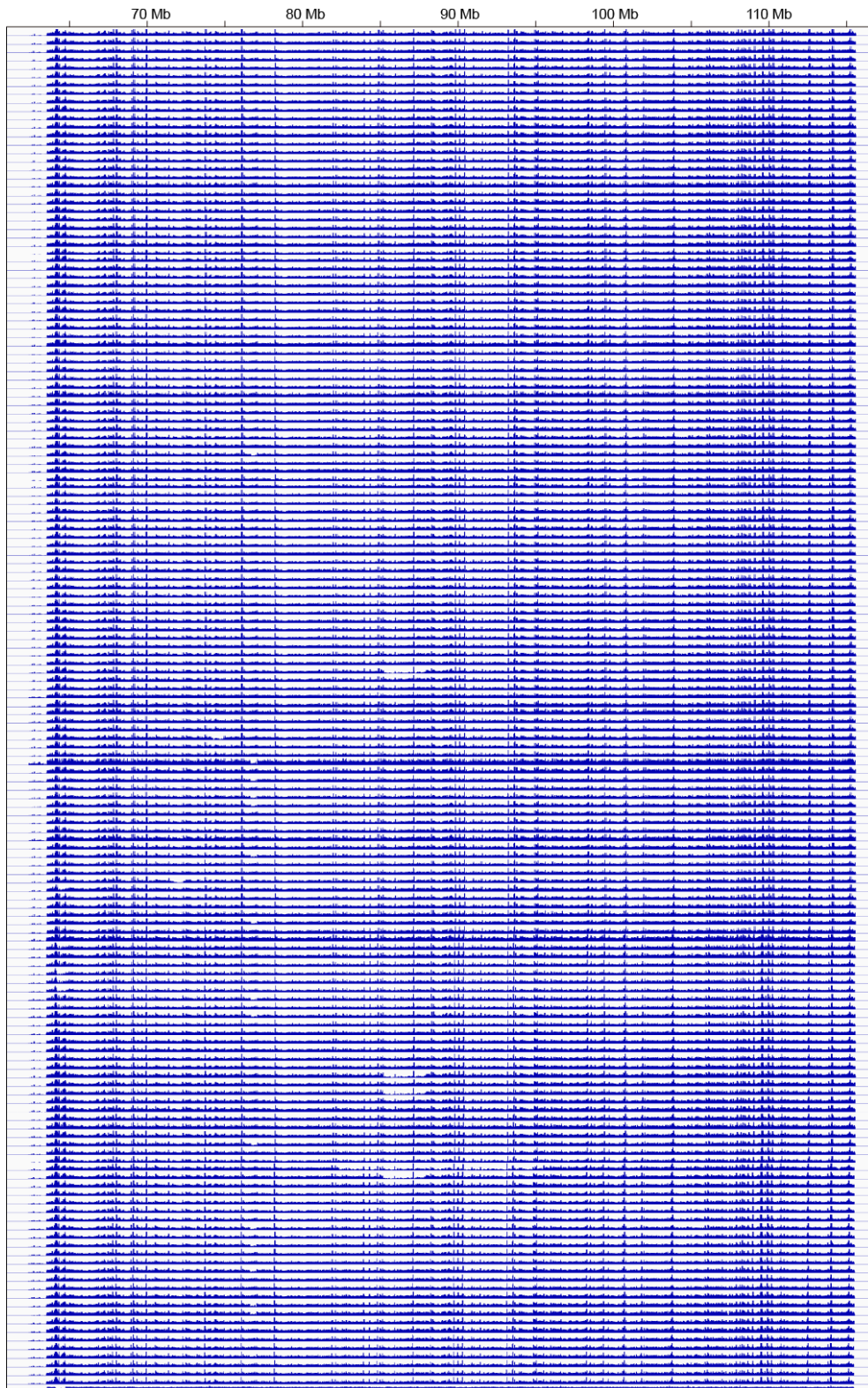
C



d

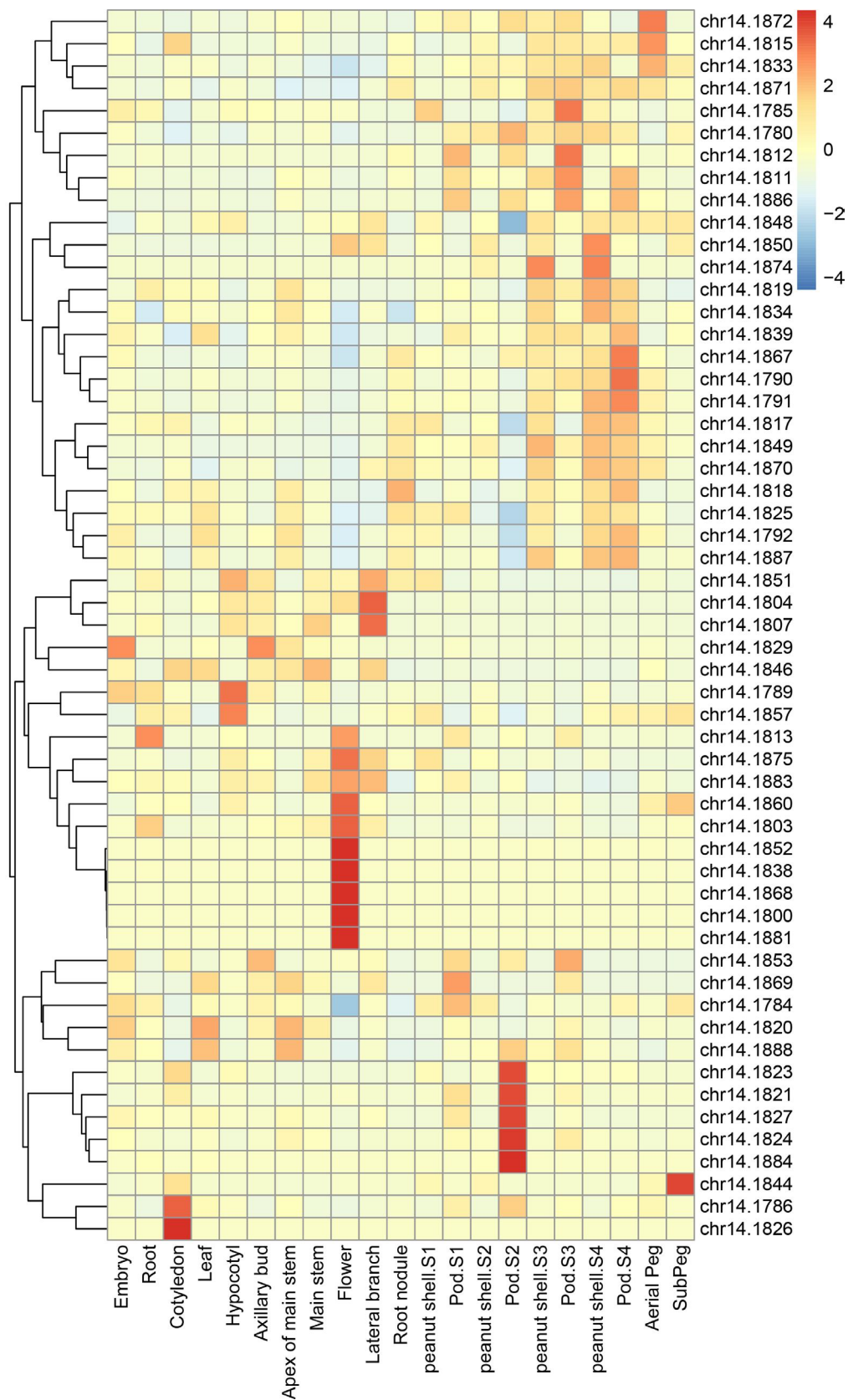


e

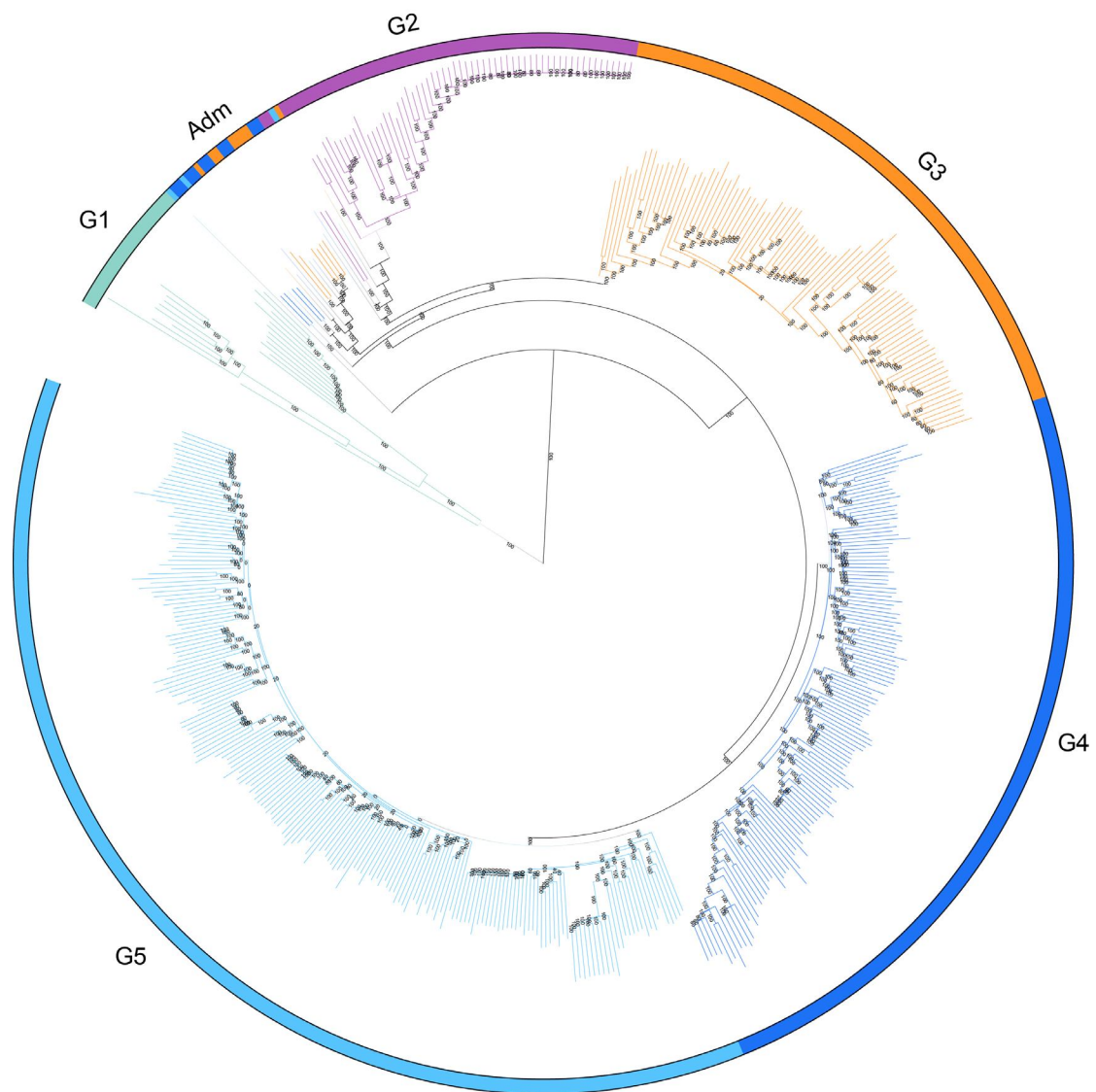


Supplementary Fig. 11 Structural variation (SV) among different botanical varieties.

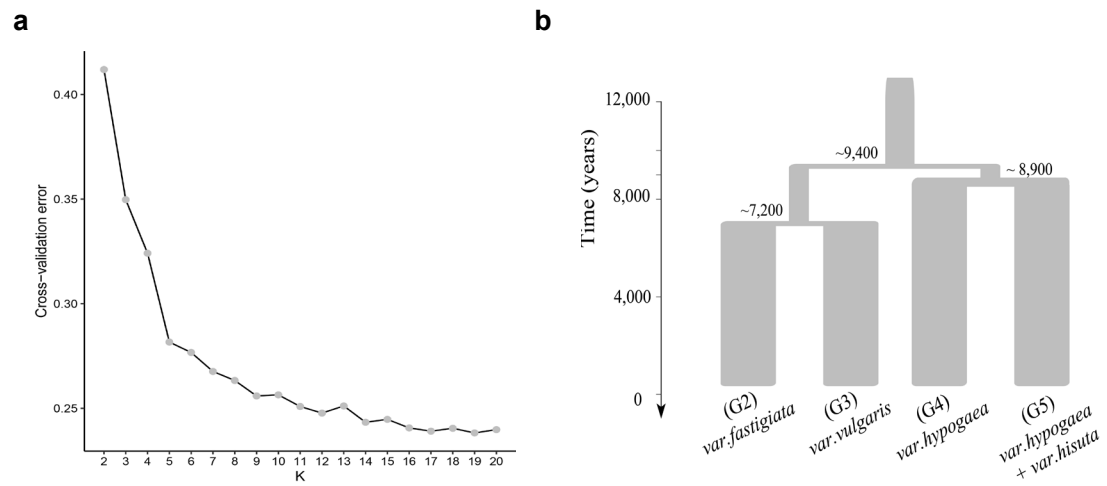
The coverage of 13.23 Mb SVs regions on Chr14, with K30076 as reference genome. **a**, wild diploid, **b**, *var. fastigiata*; **c**, *var. vulgaris*; **d**, *var. hypogaea*; **e**, *var. hypogaea* + *var. hirsuta*. Presence and absence are indicated by the coverage depth.



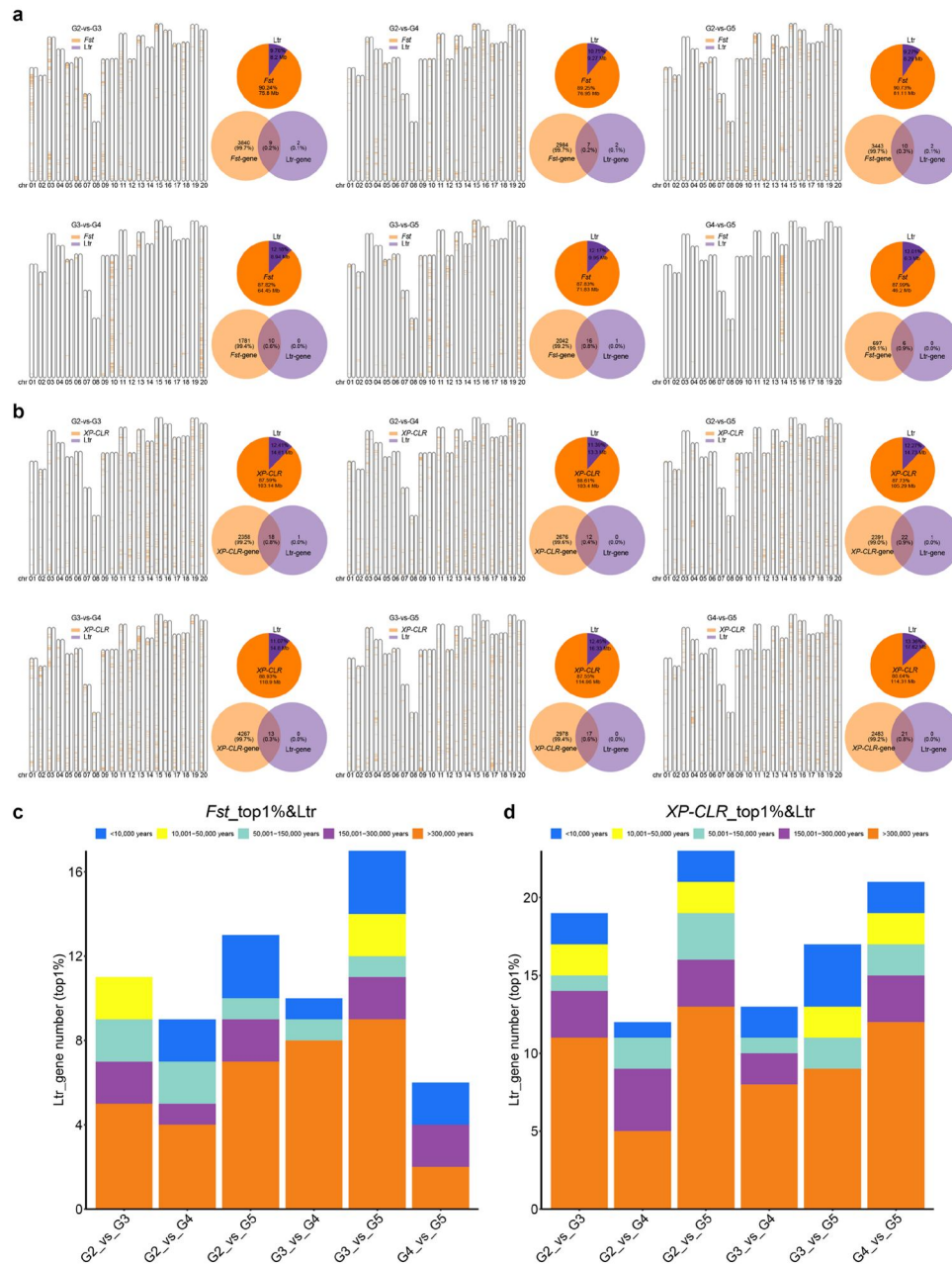
Supplementary Fig. 12 The expression pattern of structural variation (SV) genes present on Chromosome 14 in the S83 accession across different tissues.



Supplementary Fig. 13 Phylogenetic tree of 521 peanut accessions. The neighbor-joining tree, with support values, divides the 521 peanut accessions into five main groups: wild diploid (G1), *var. fastigiata* (G2), *var. vulgaris* (G3), *var. hypogaea* (G4) and *var. hypogaea* + *var. hirsuta* (G5).

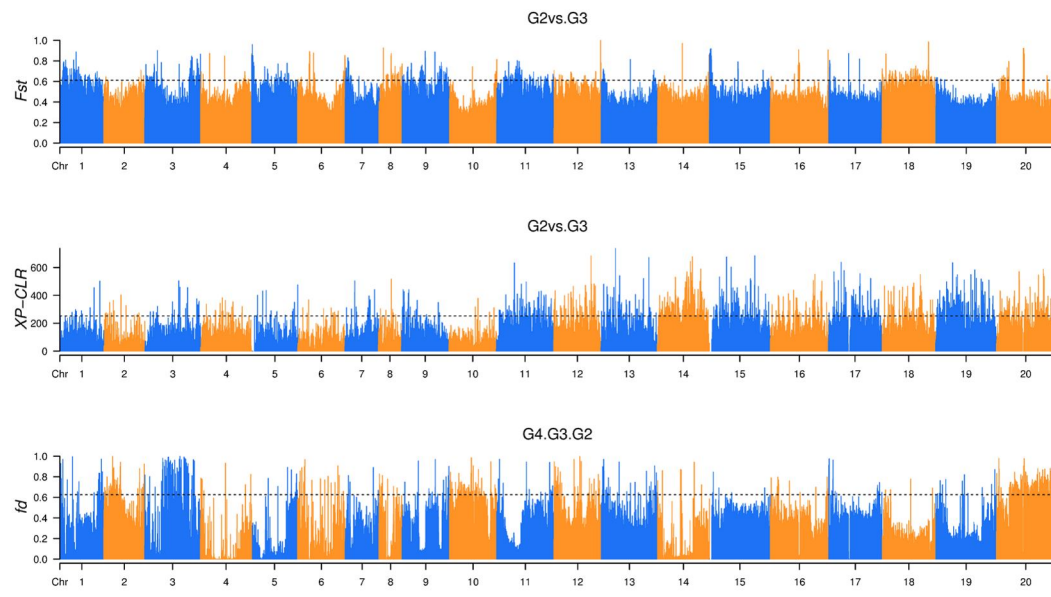


Supplementary Fig. 14 Population structure and demographic history of peanut populations. **a**, Cross-validation error values for different K values (K = 2 to 20). **b**, Schematic representation of divergence times estimated using fastsimcoal2. The vertical axis indicates the estimated timing of population divergence.

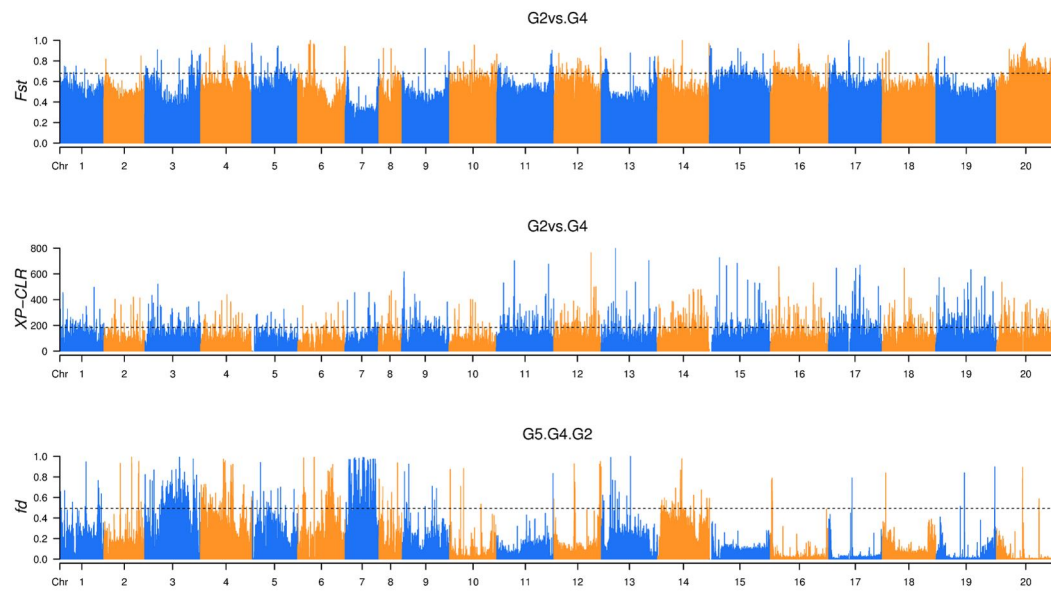


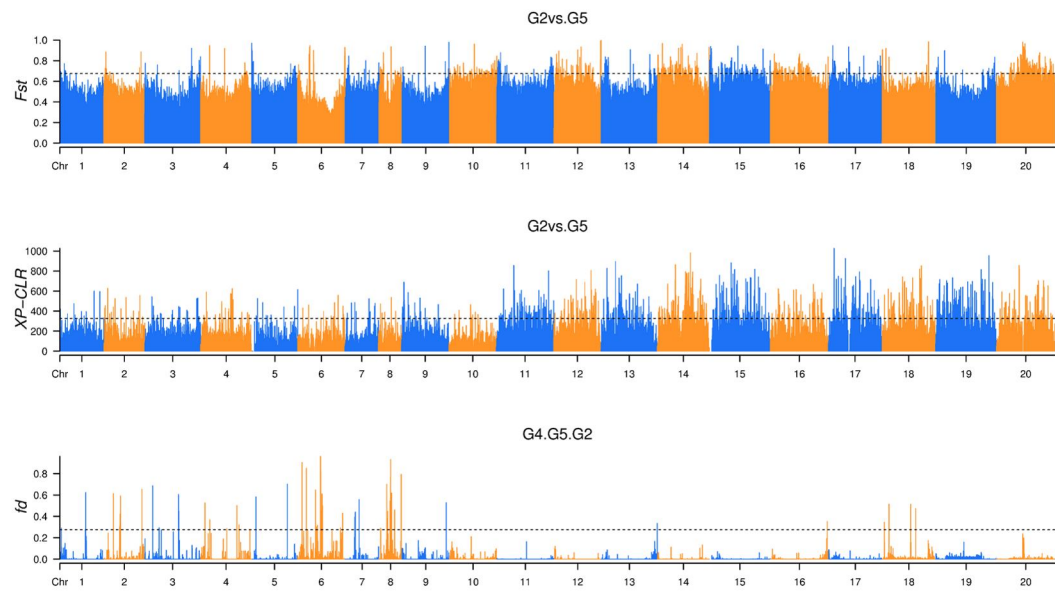
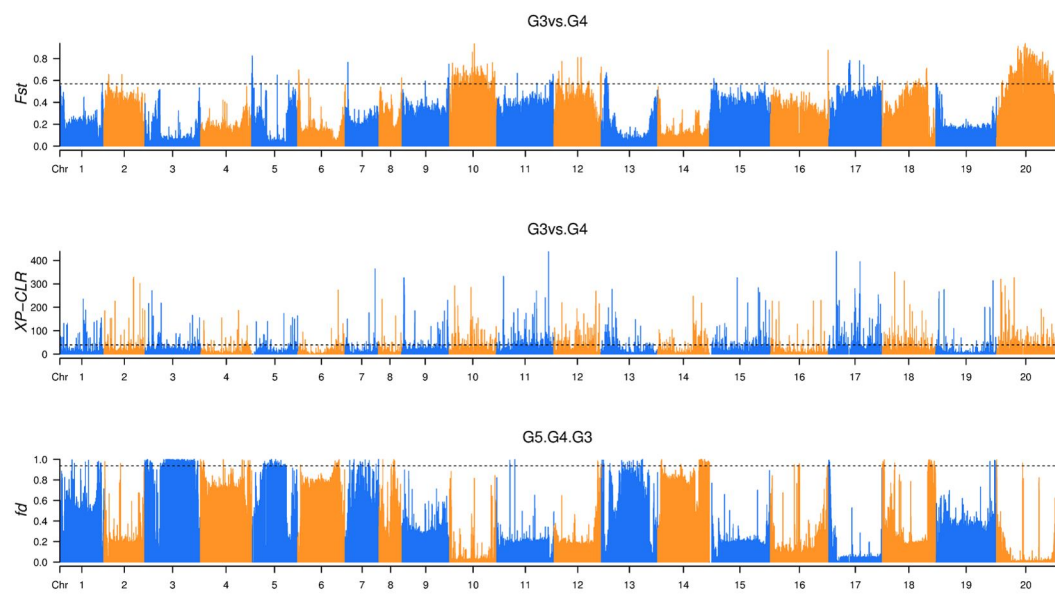
Supplementary Fig. 15 LTR insertion on selected sweep regions. The selective sweep analysis among six different populations was conducted using *Fst* and *XP-CLR*. Selective sweep regions (left) and selective sweep regions with LTR insertion regions (right) were plotted on the chromosome schematic in the left of Fig. **a** and **b**. The proportion of selective sweep intervals and LTRs, along with the corresponding number of genes, is presented as pie charts in the left of Fig A and B. LTR insertions are categorized into five time periods based on insertion times. *Fst_top1%* (**c**) and *XP-CLR_top1%* (**d**) refer to the top 1% intervals retained, respectively. LTR-genes are defined as genes that overlap with LTR elements.

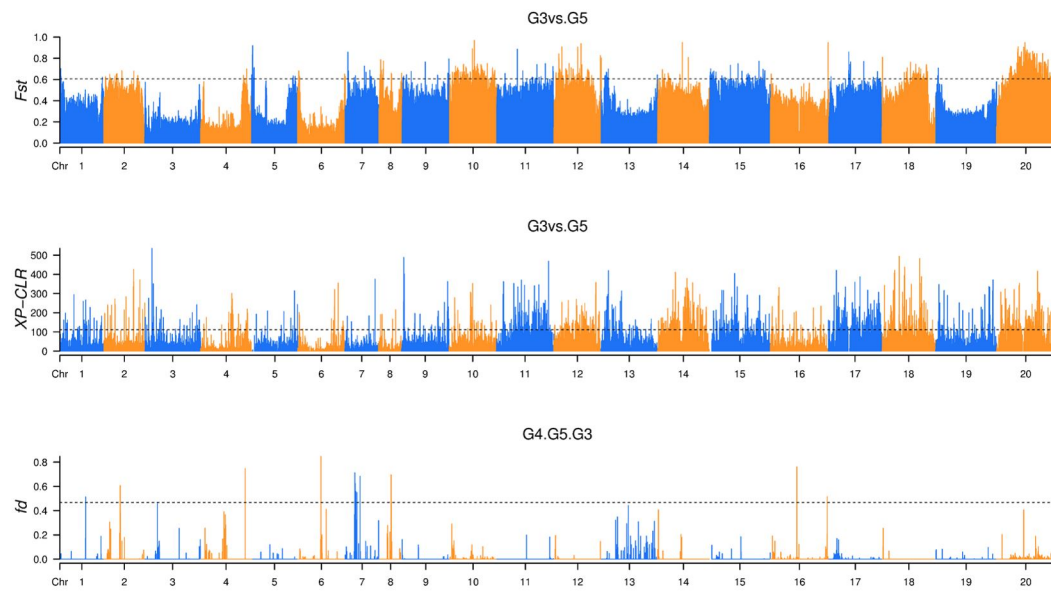
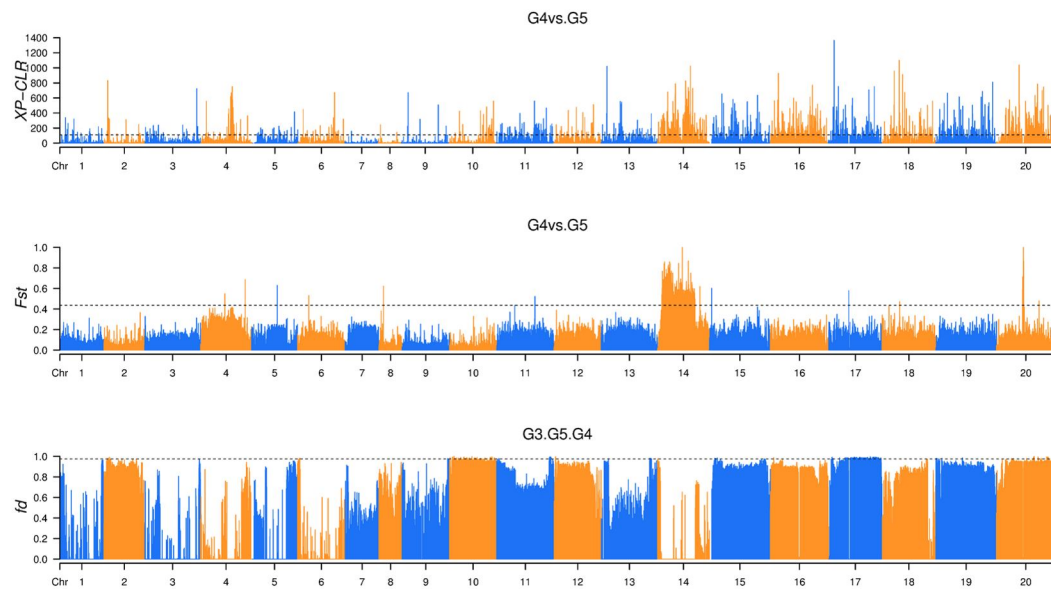
a



b

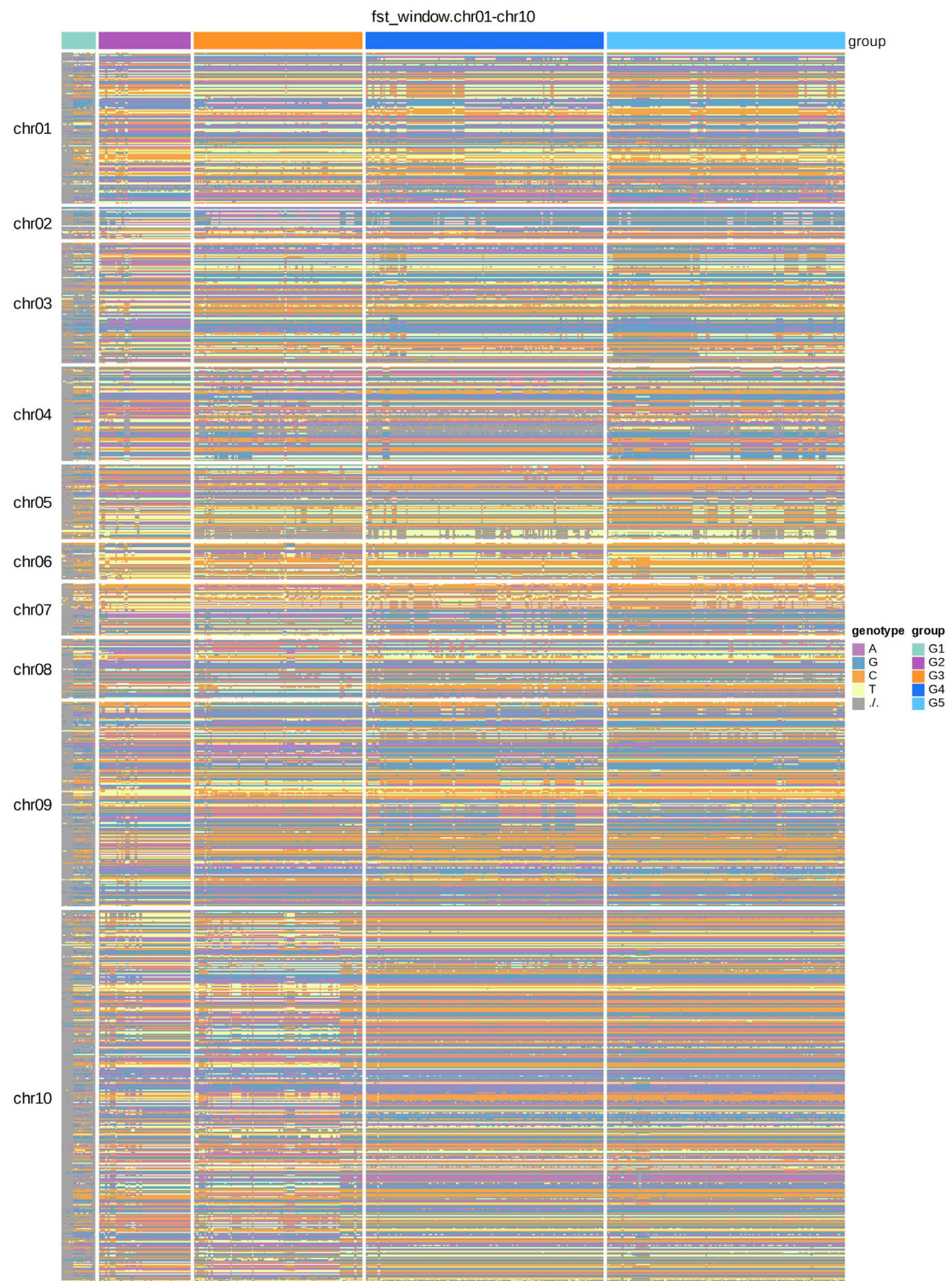


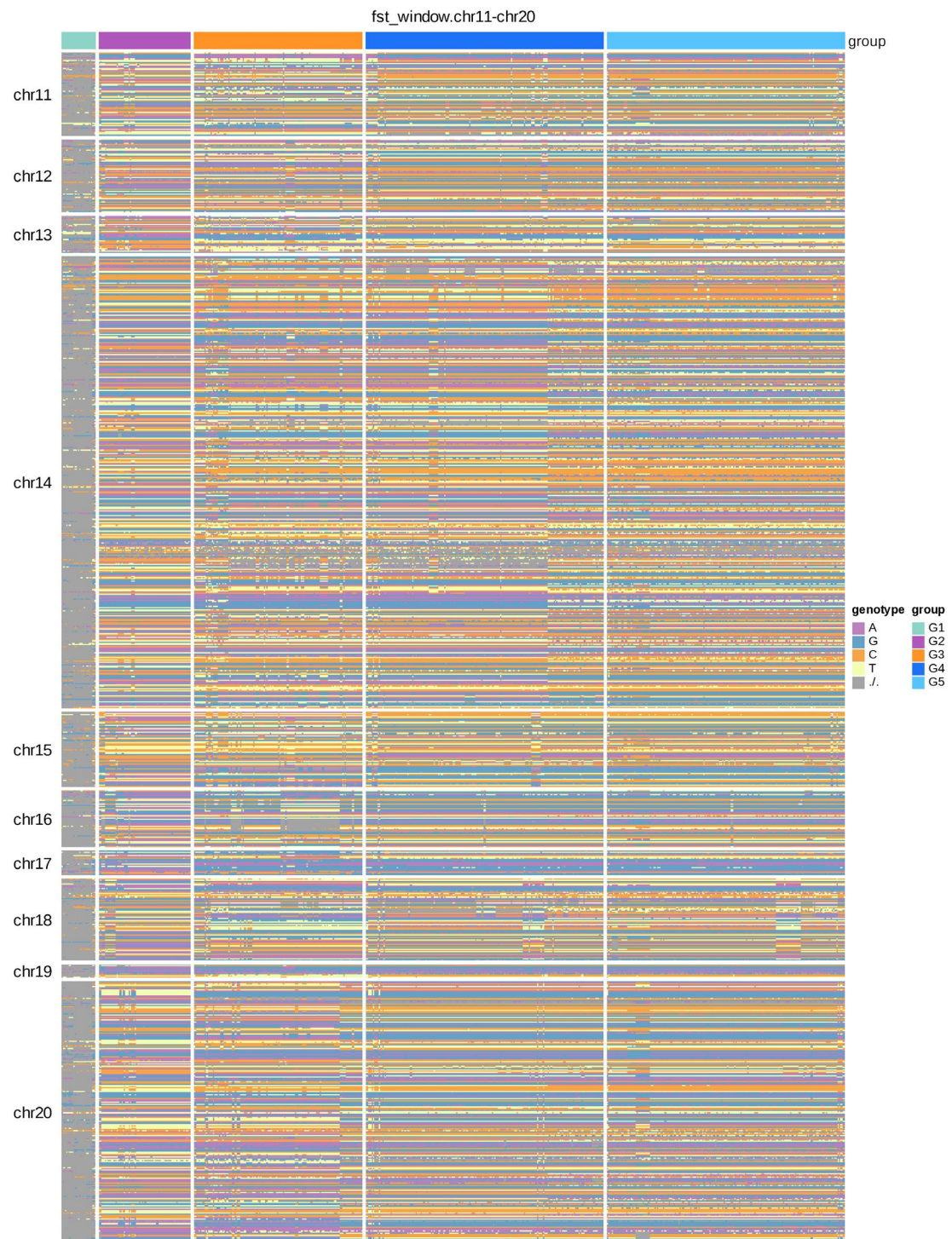
c**d**

e**f**

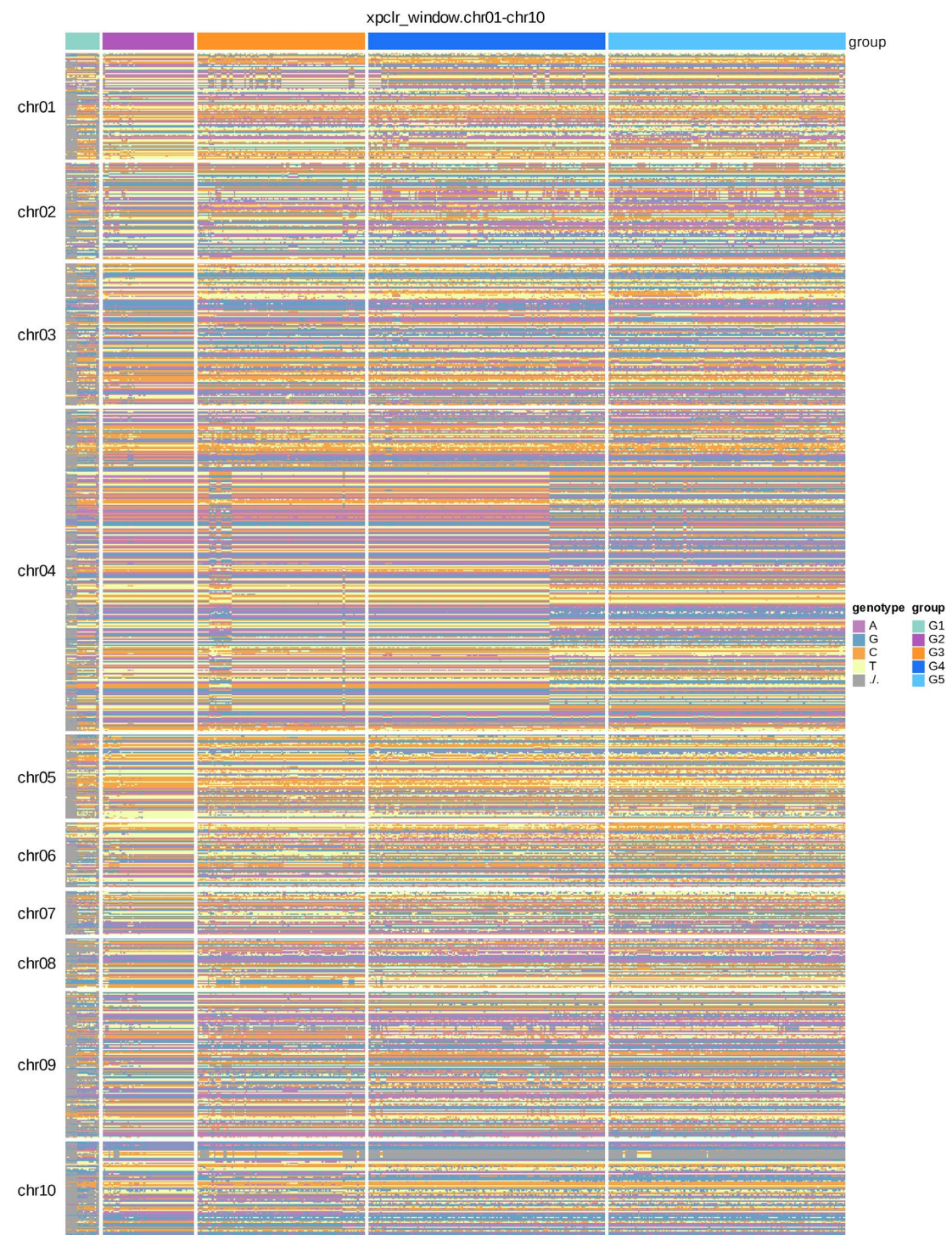
Supplementary Fig. 16 Selected sweep and introgression among different populations. F_{st} , $XP-CLR$, and f_d analysis were conducted across different chromosomes among the following groups: **a**, G2-G3; **b**, G2-G4; **c**, G2-G5; **d**, G3-G4; **e**, G3-G5; **f**, G4-G5. The threshold line is set at the top 1%.

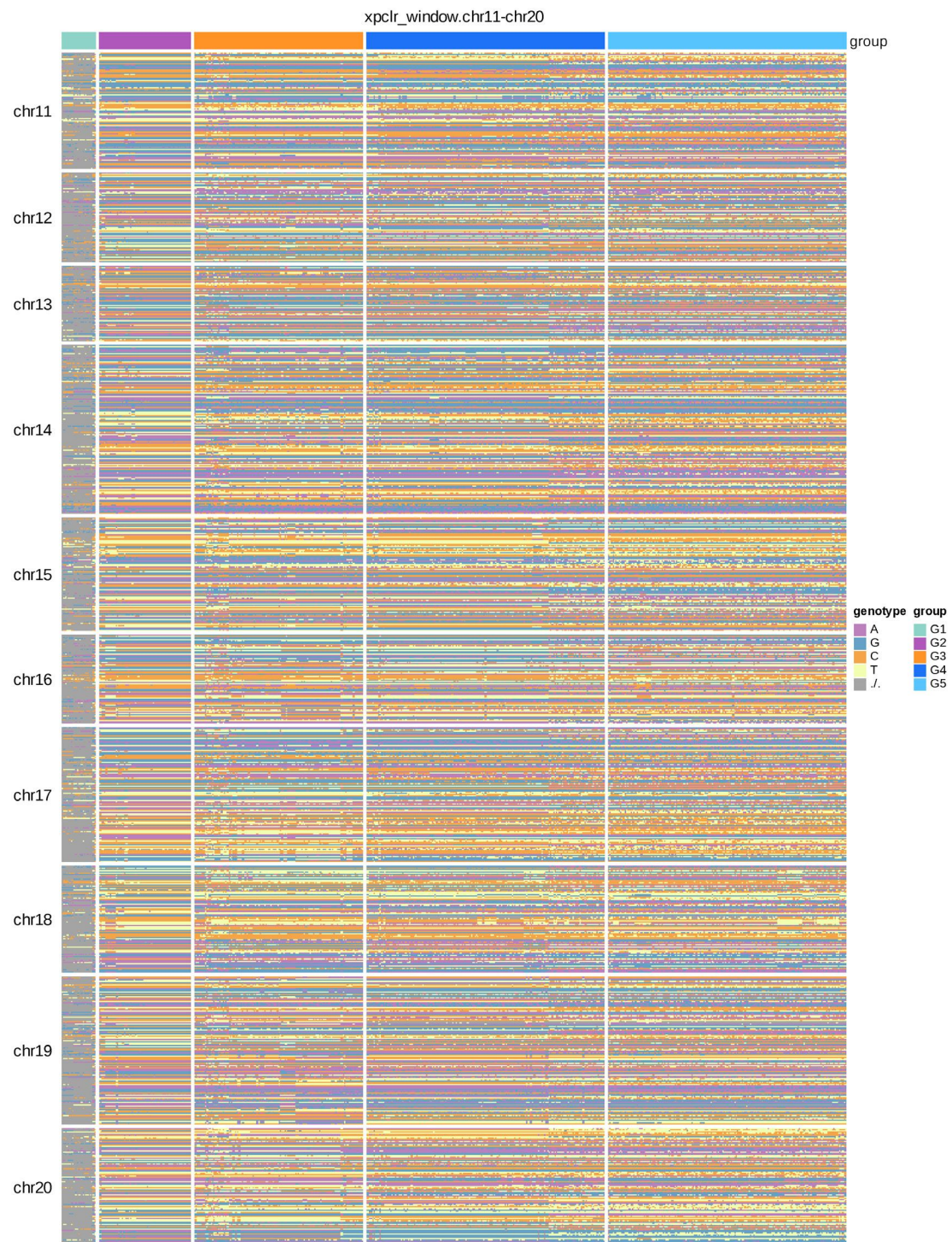
a



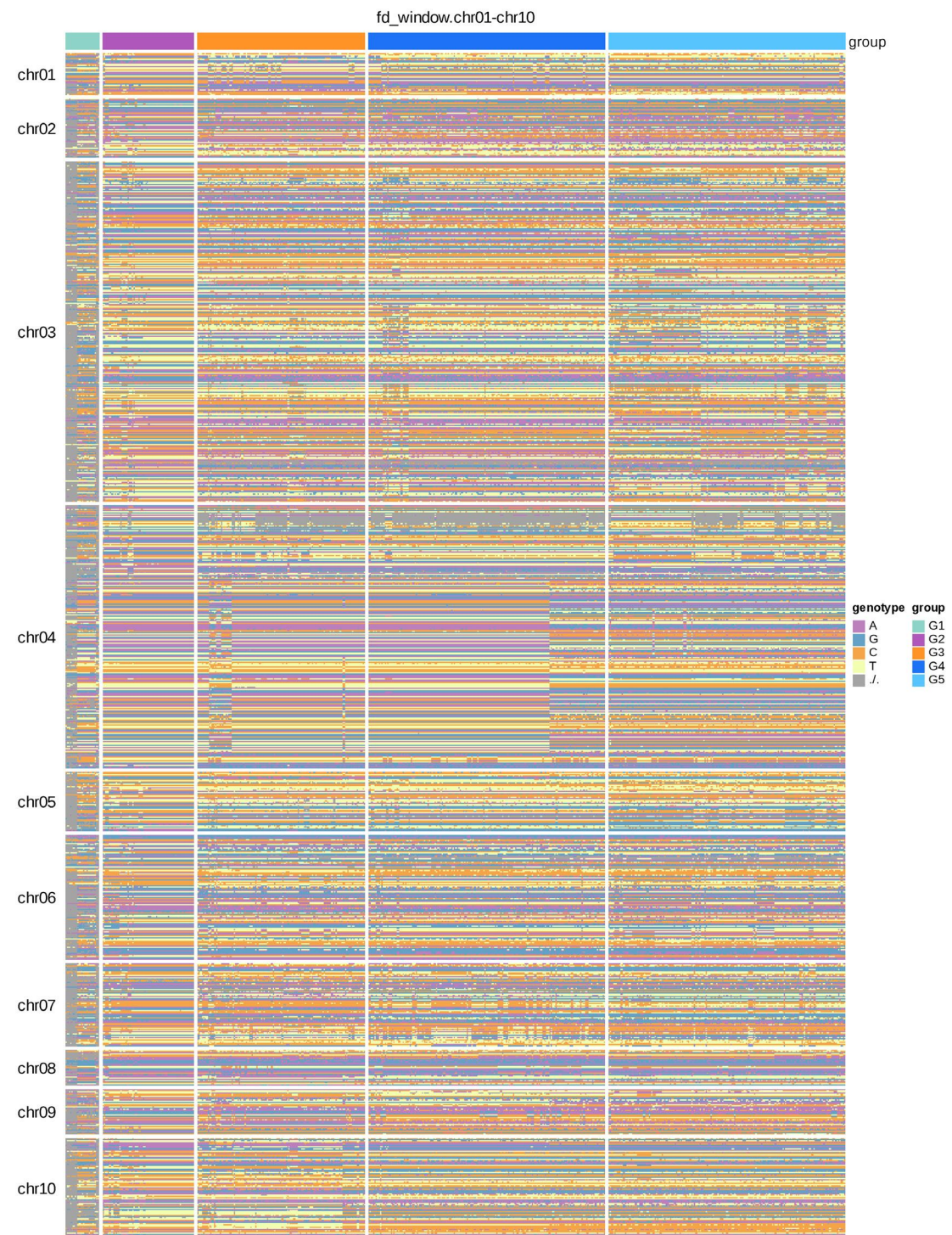


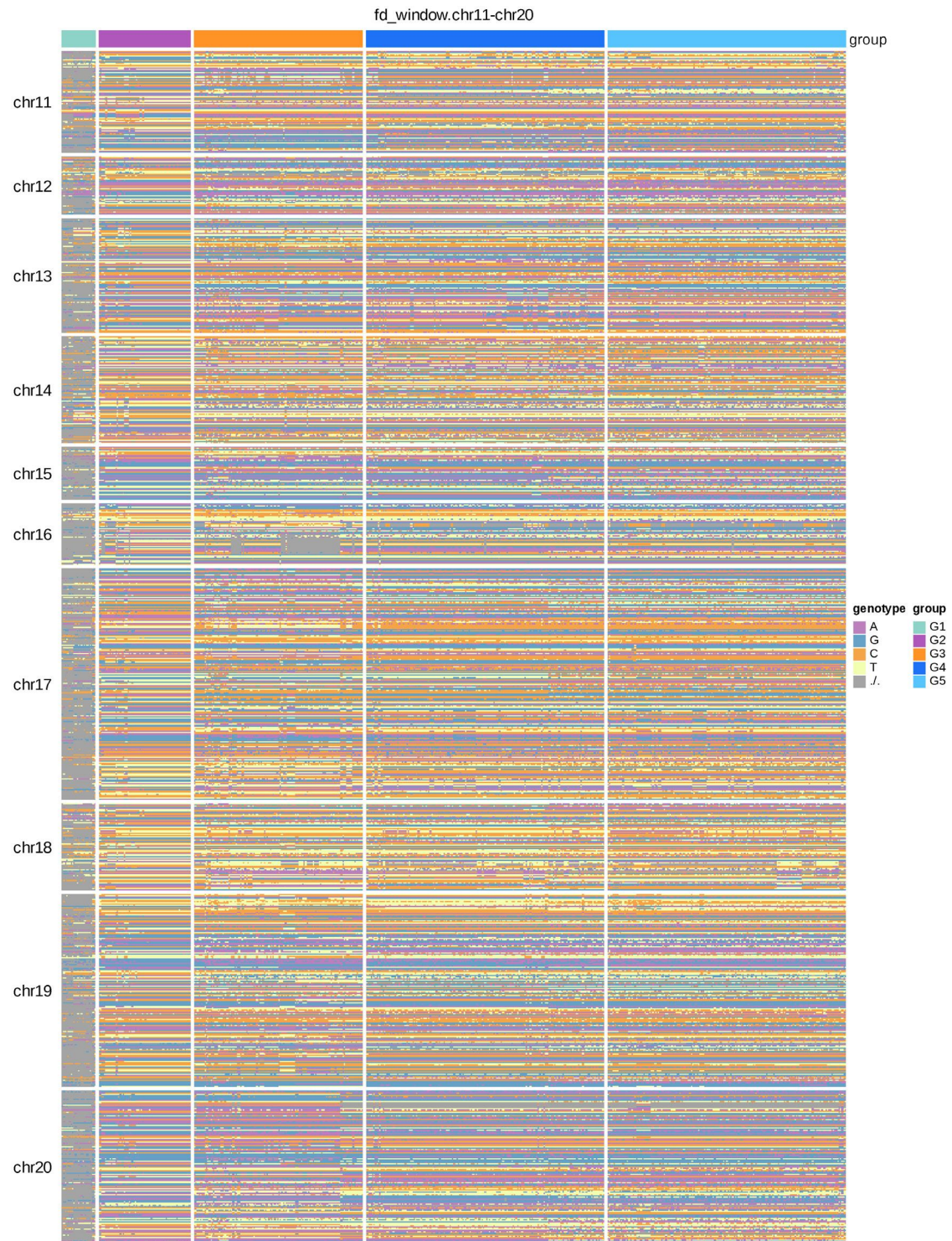
b





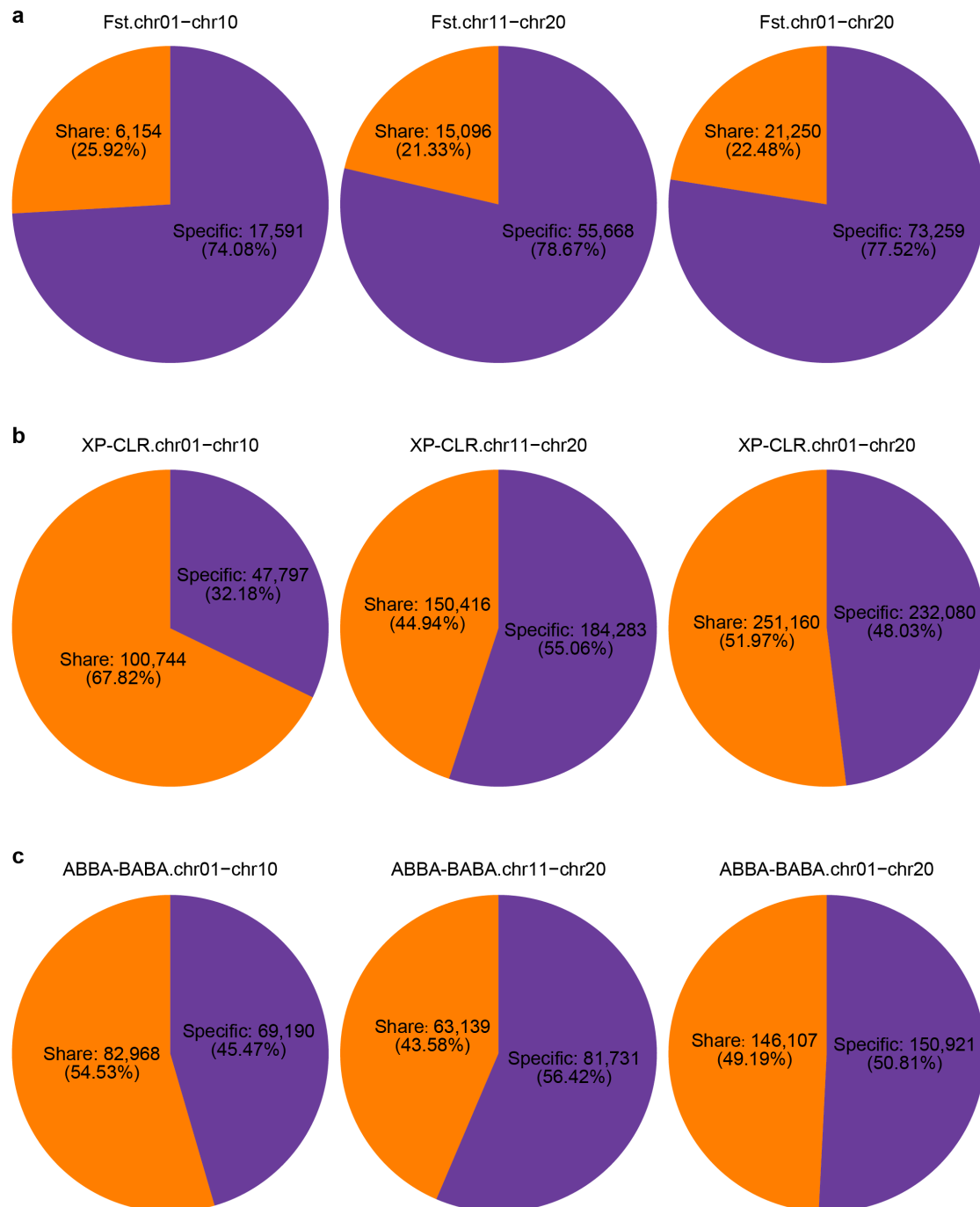
c





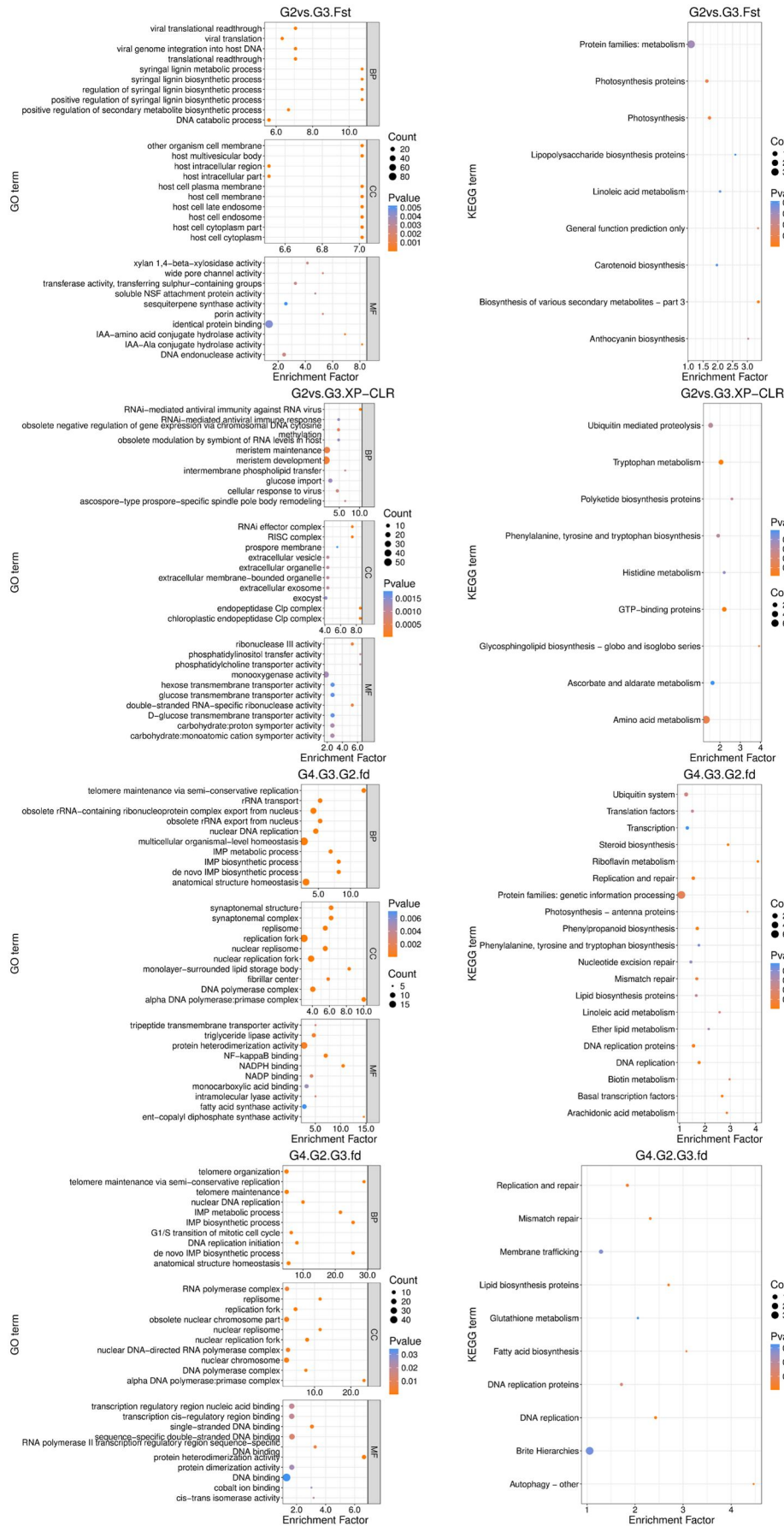
Supplementary Fig. 17 SNP Variation Types in Selected and Introgressed Regions.

The distribution of SNP variants (A, G, C, T, and ./.) across Group1 to Group5 is visualized with different colors. The S245 reference genome is used for comparison. **a**, Top 1% *Fst* regions. **b**, Top 1% *XP-CLR* regions. **c**, Top 1% Introgressed regions.

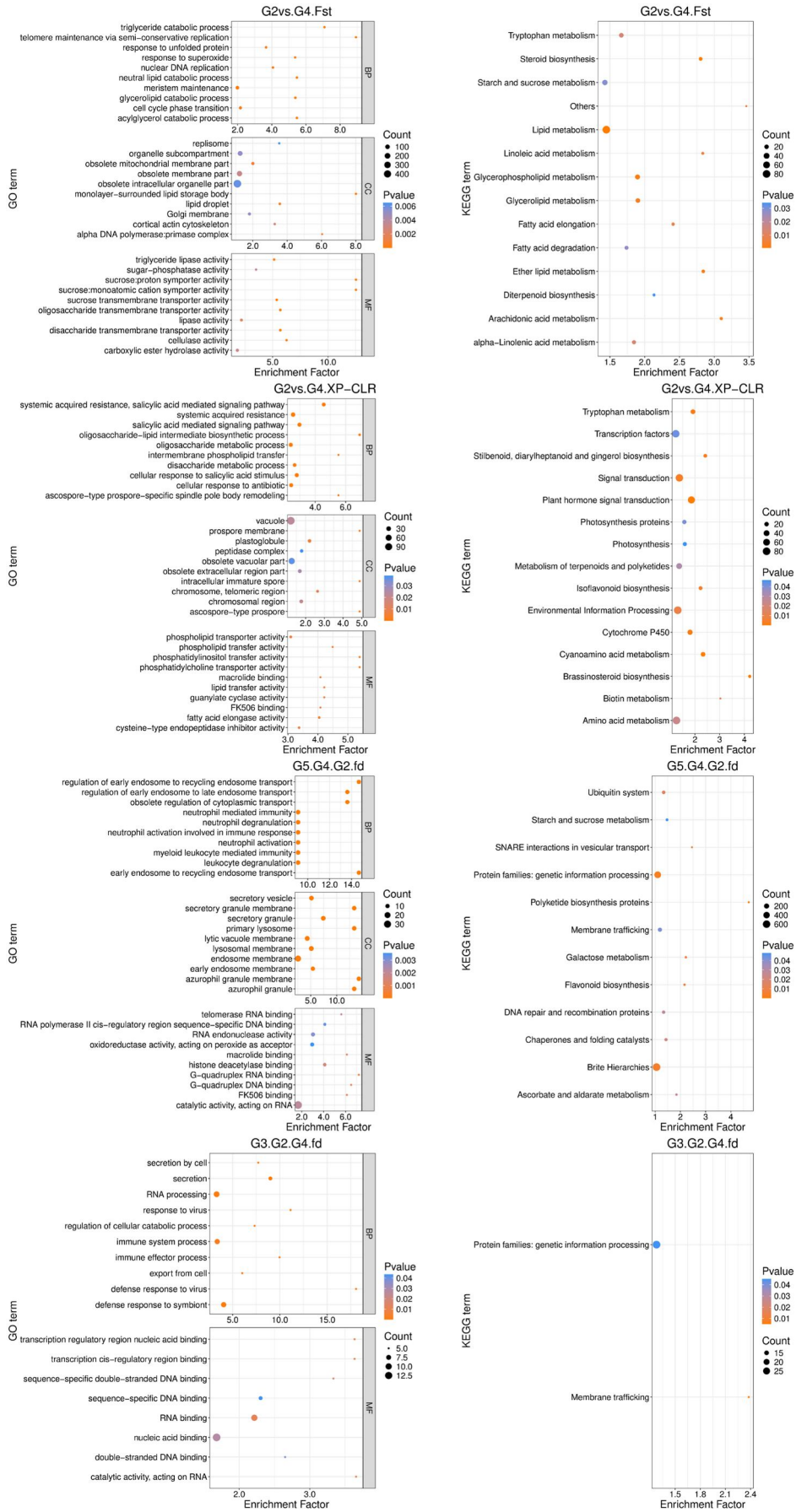


Supplementary Fig. 18 Percentage of share and specific variation between diploids and tetraploids. The variation type in tetraploids is defined as shared variation if it exists in the wild accessions; otherwise, it is considered tetraploid-specific variation. **a**, Top 1% *Fst* regions. **b**, Top 1% *XP-CLR* regions. **c**, Top 1% introgressed regions. From left to right, represent the At (chr01–chr10), Bt (chr11–chr20), and At Bt (chr01–chr20) subgenomes.

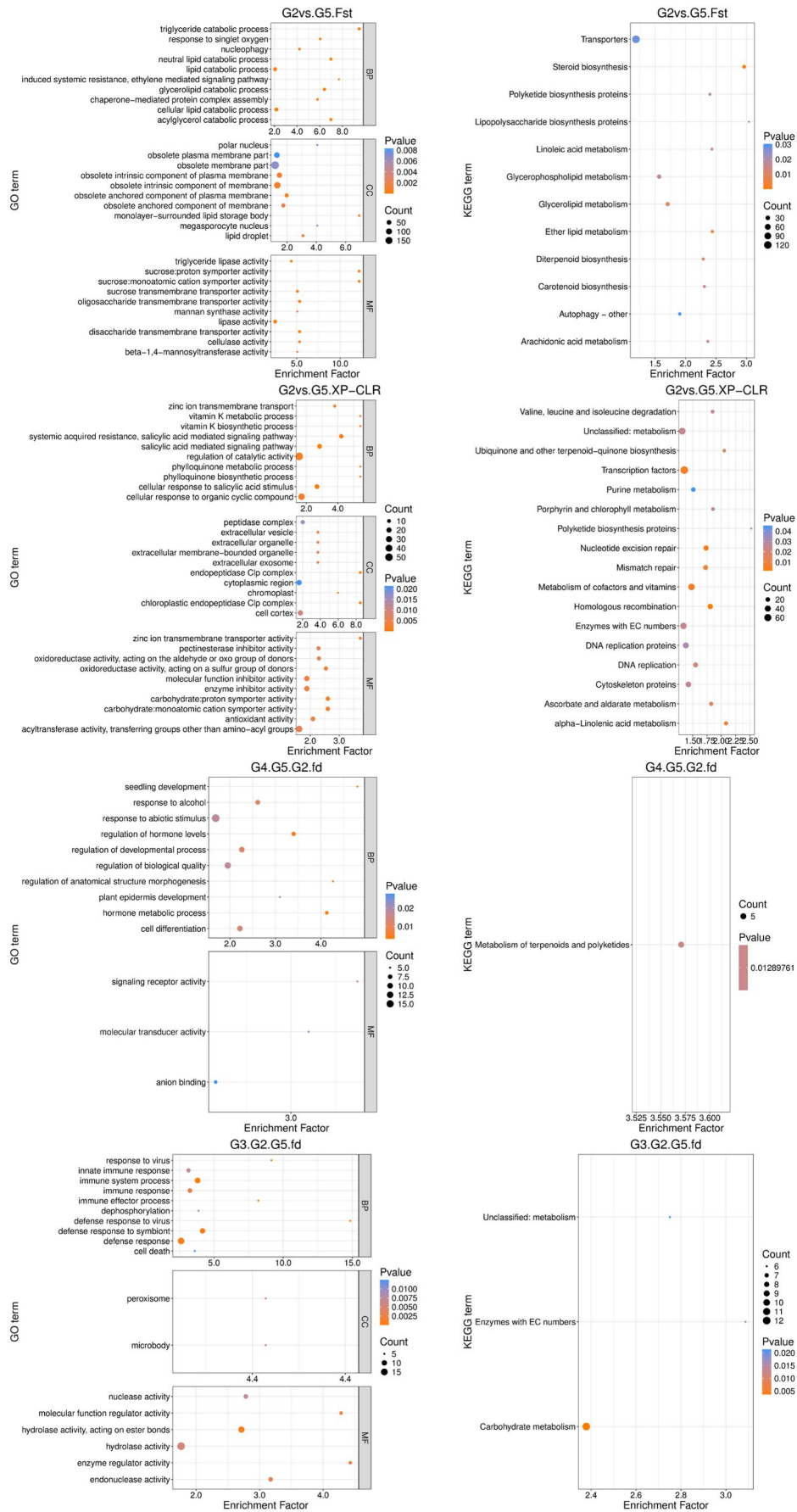
a



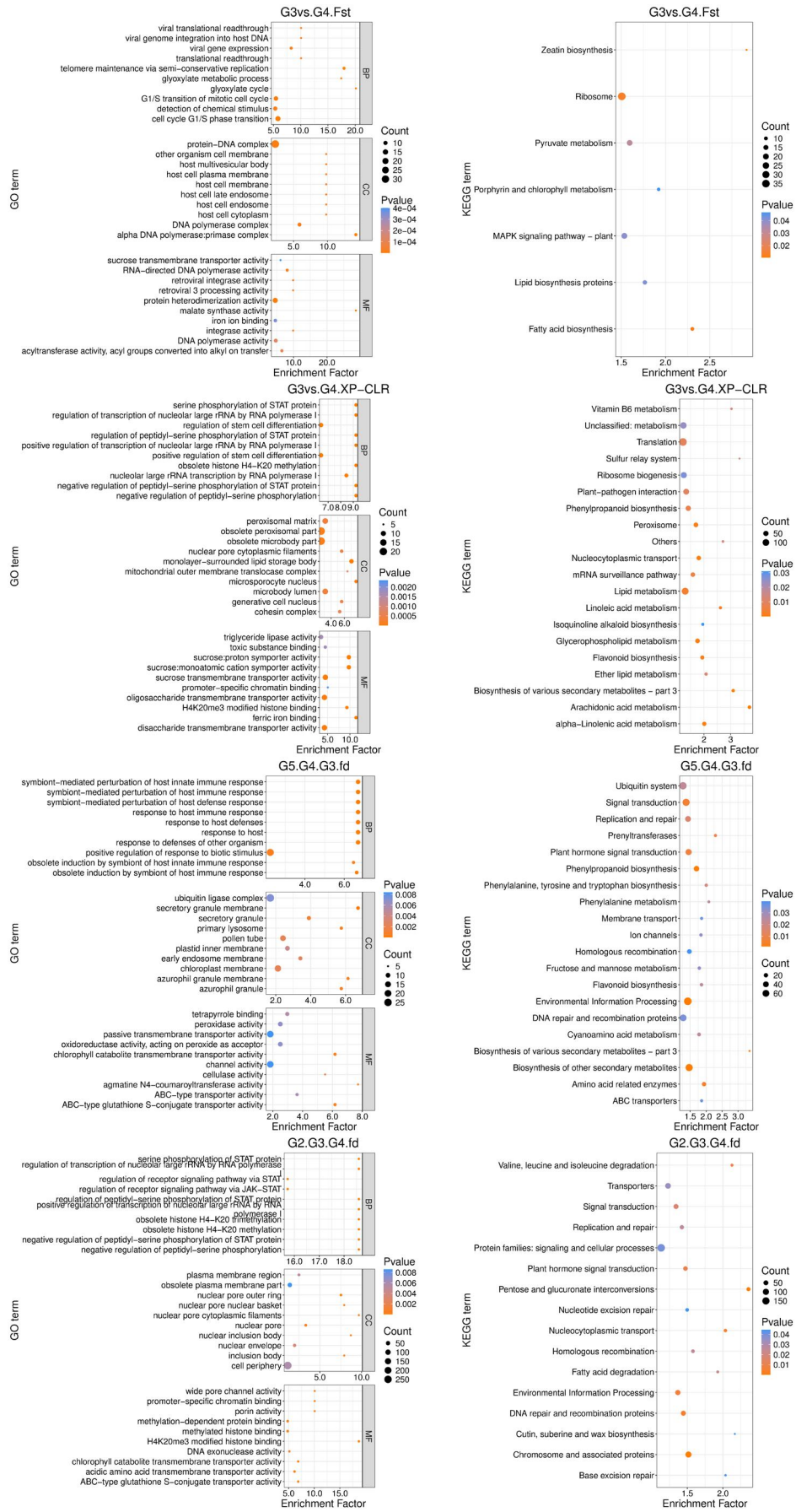
b



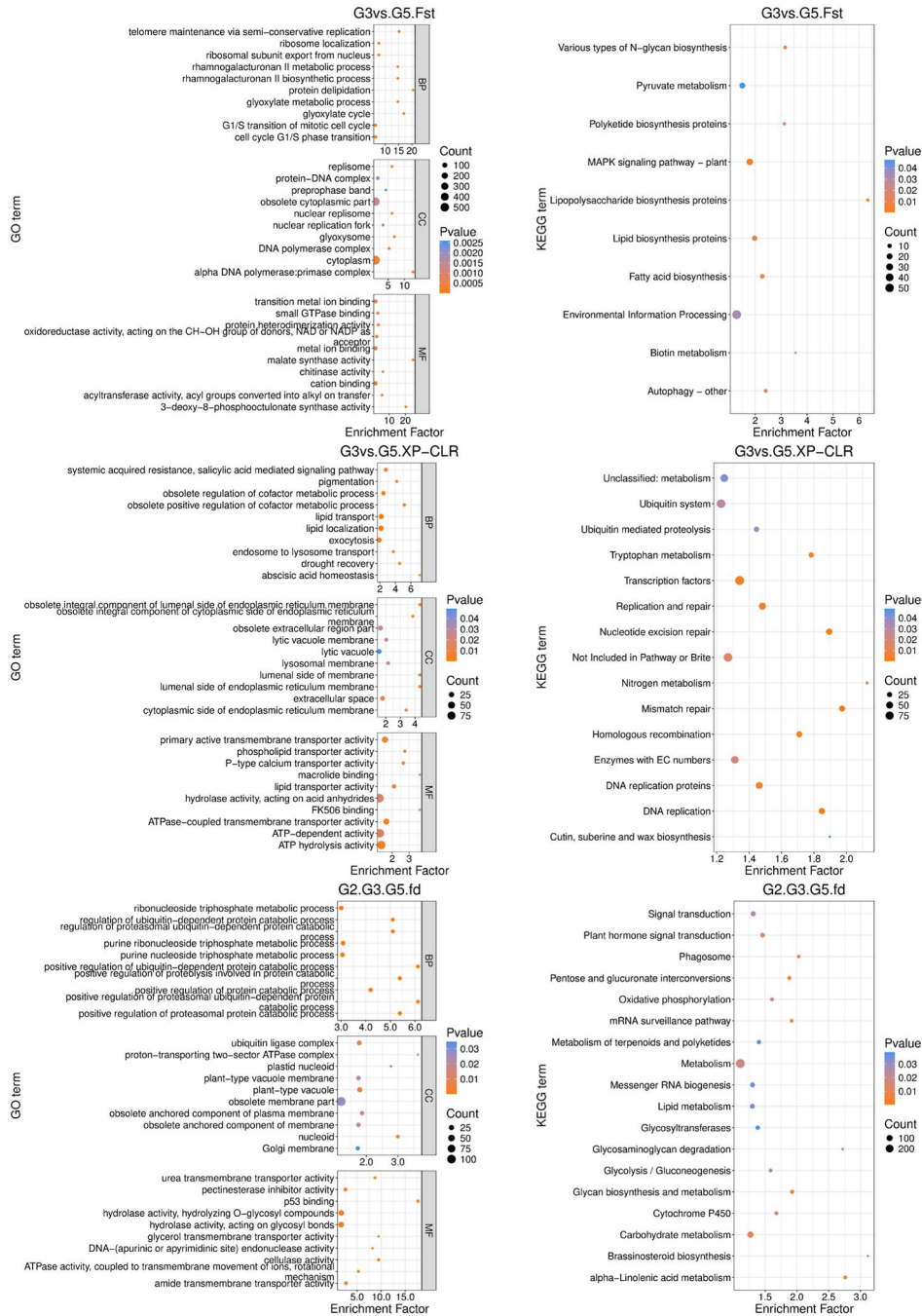
C



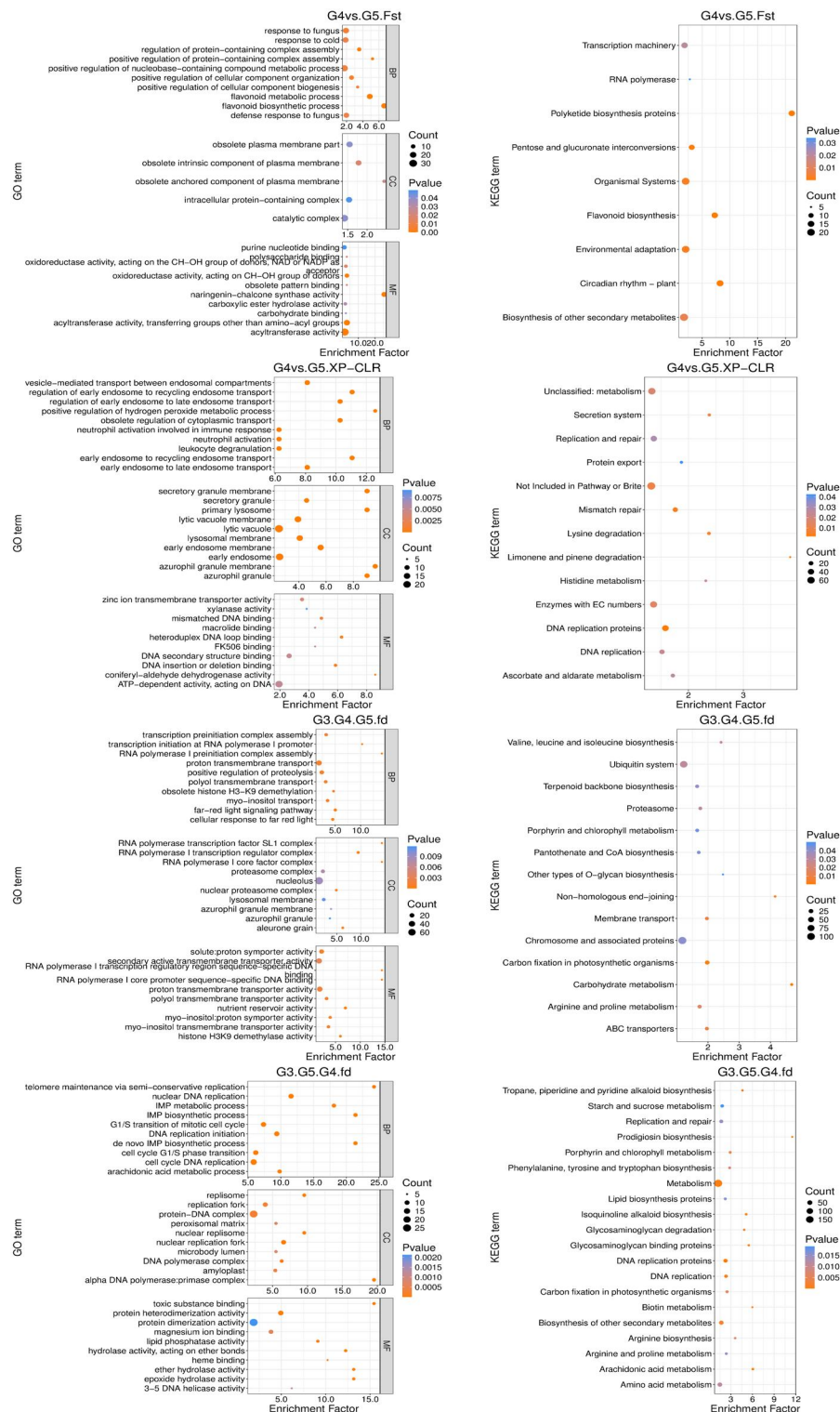
d



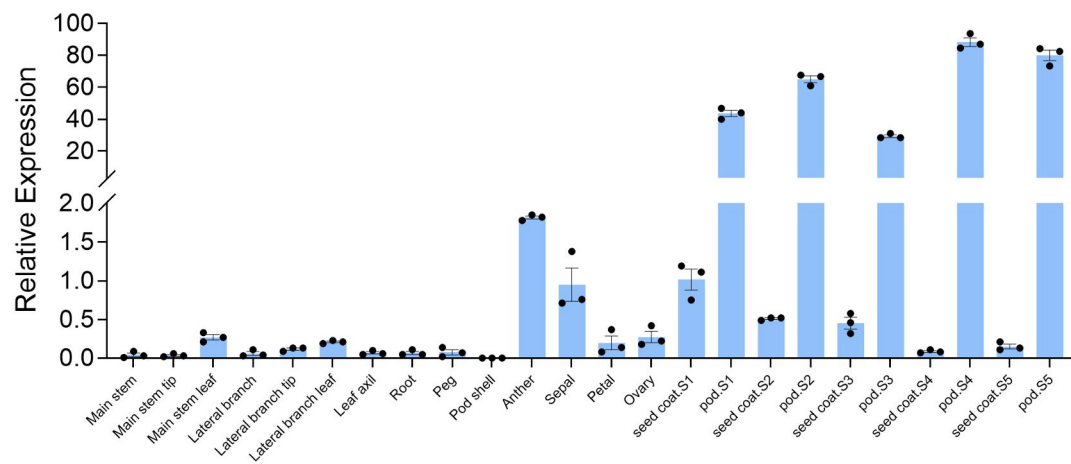
e



f

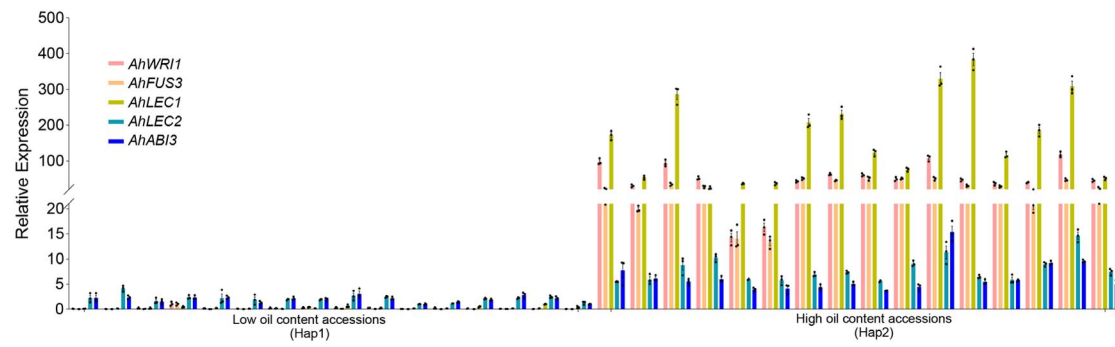


Supplementary Fig. 19 Enrichment analyses of significantly selected and introgression among different populations. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment were conducted for genes located in the top 1% of selected and introgression regions among the following groups: **a**, G2-G3; **b**, G2-G4; **c**, G2-G5; **d**, G3-G4; **e**, G3-G5, and **f**, G4-G5. *p-value* < 0.05.

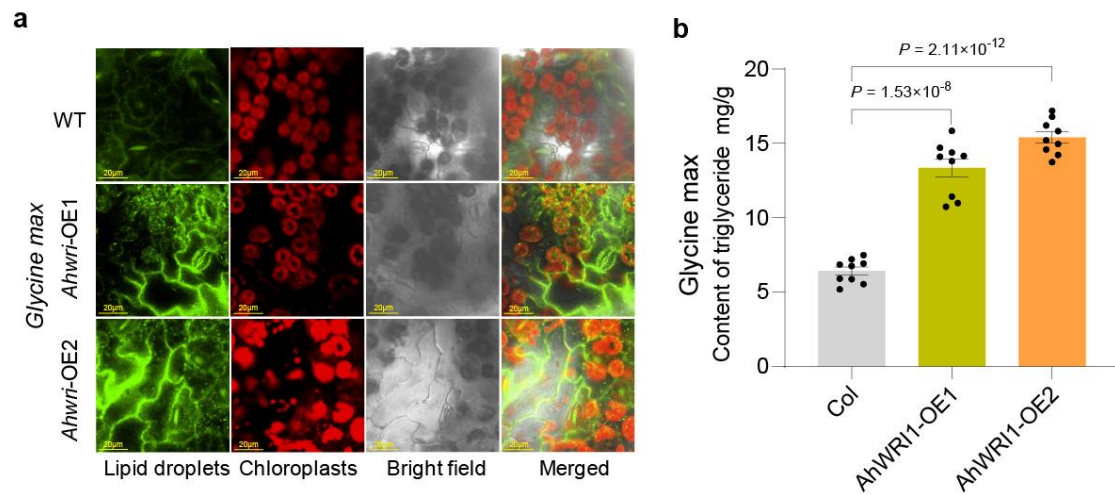


Supplementary Fig. 20 qRTPCR result of tissue-specific expression of *AhWR1* gene.

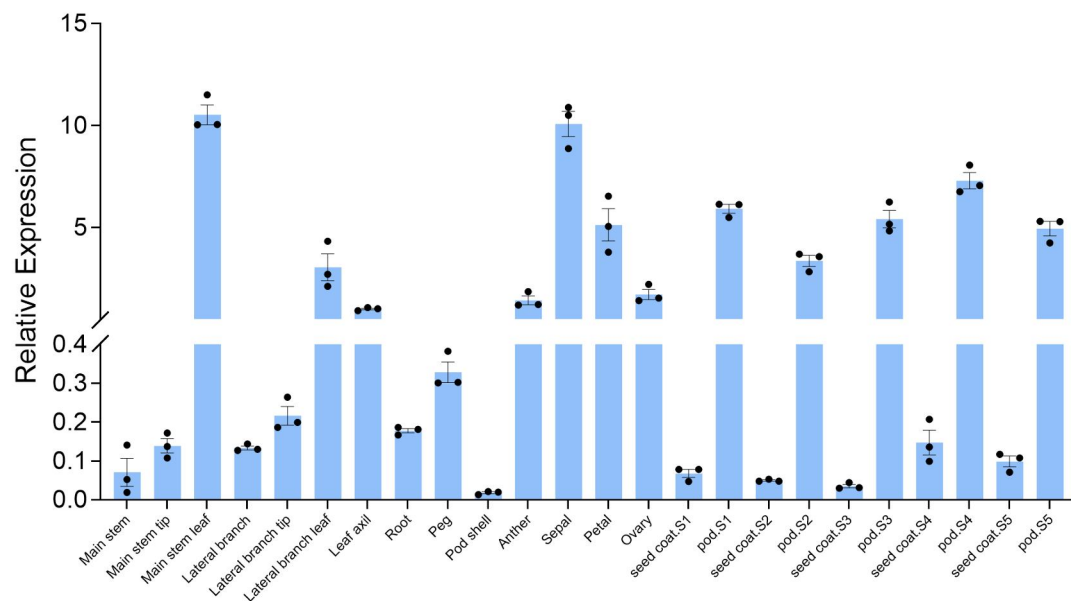
S1-S5: 15, 30, 45, 60, 75 days after flowering. Quantitative data are mean $\pm$ s.e.m. n = 3 biologically independent samples.



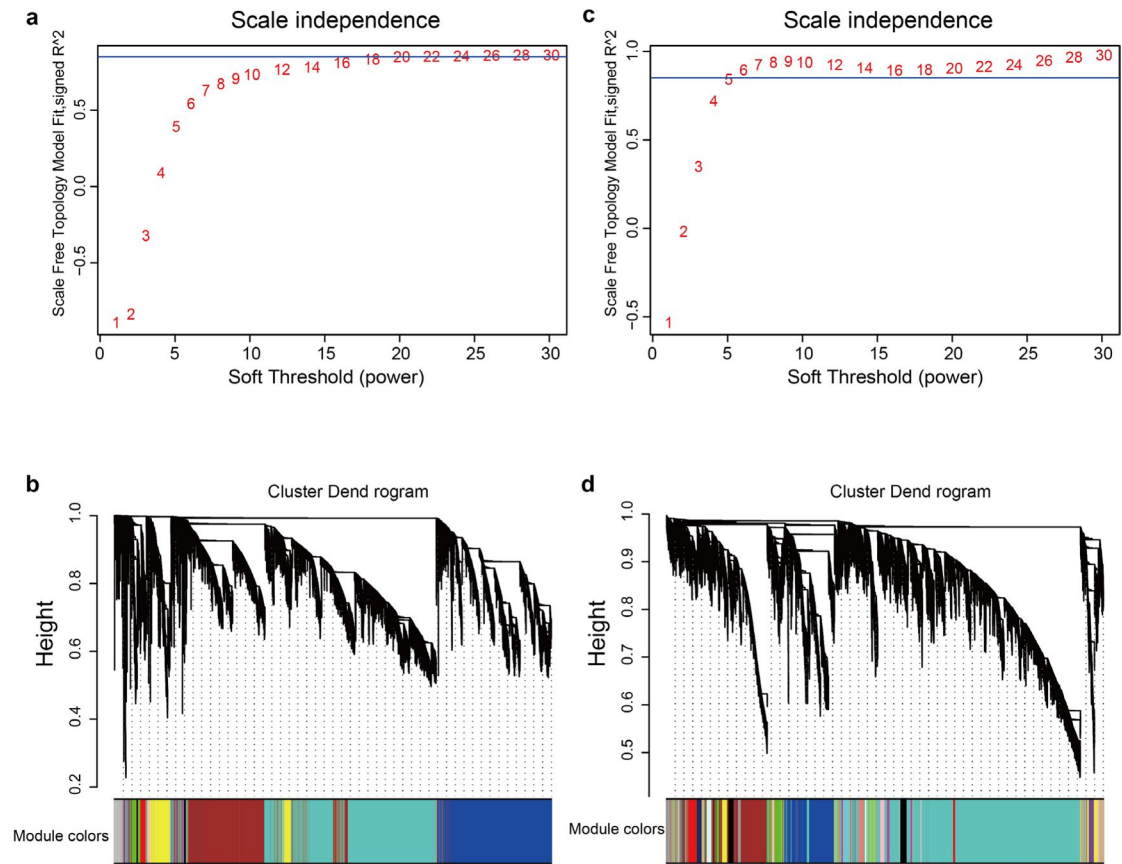
Supplementary Fig. 21 The expression levels of *AhFUS3*, *AhABI3*, *AhLECs*, and their target gene *AhWR11* in high and low oil content peanut accessions. Quantitative data are mean $\pm$ s.e.m. n = 3 biologically independent samples.



Supplementary Fig. 22 The oil content identification of overexpressing *AhWR11* in *Glycine max* plant. **a**, Confocal images of accumulated lipid droplets (LDs) stained with Nile Red (green) in *Glycine max* transgenic plant leaf infected with *Agrobacterium*. Scale bar corresponds to 20 μ m. **b**, Triacylglycerols content of *Glycine max* plant seeds transiently overexpressing *AhWR11* isoforms. Quantitative data are mean $\pm$ s.e.m. $n = 9$ biologically independent samples. P values were calculated by two-tailed Student's t -tests.

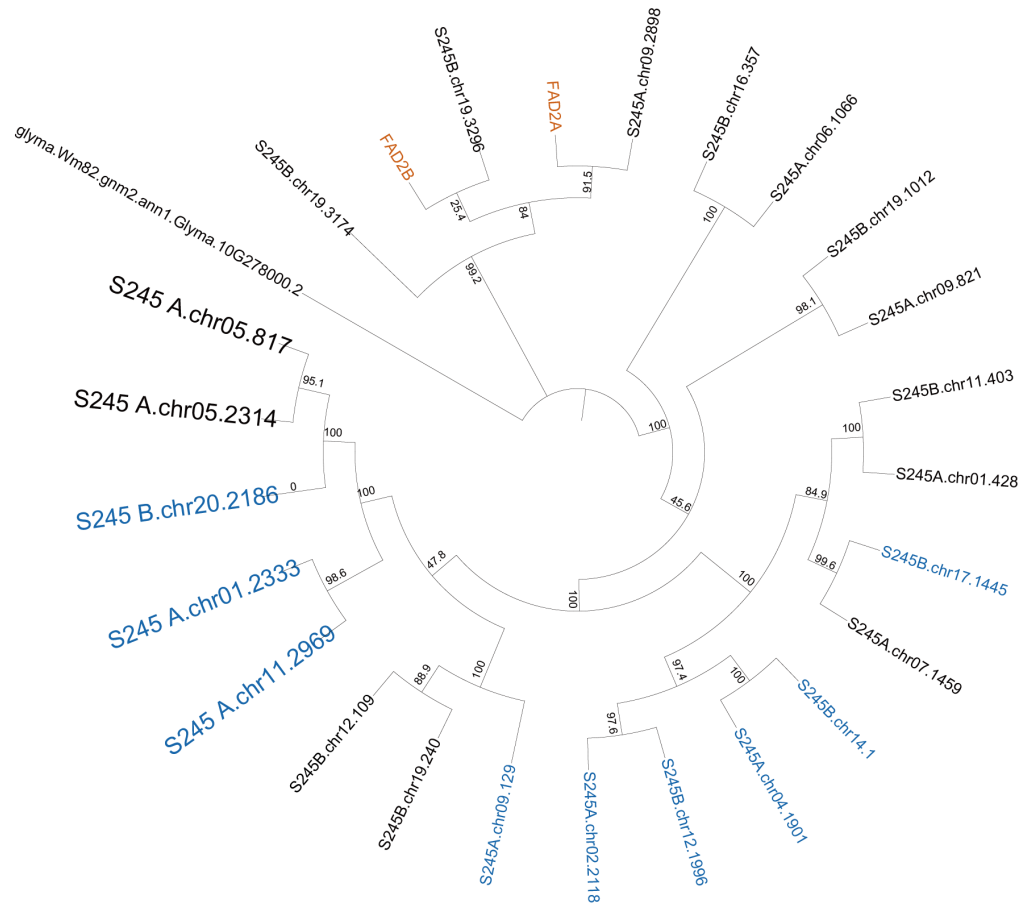


Supplementary Fig. 23 Tissue-specific expression of *AhGSA1* gene. S1-S5: 15, 30, 45, 60, 75 days after flowering. Quantitative data are mean $\pm$ s.e.m. n = 3 biologically independent samples.

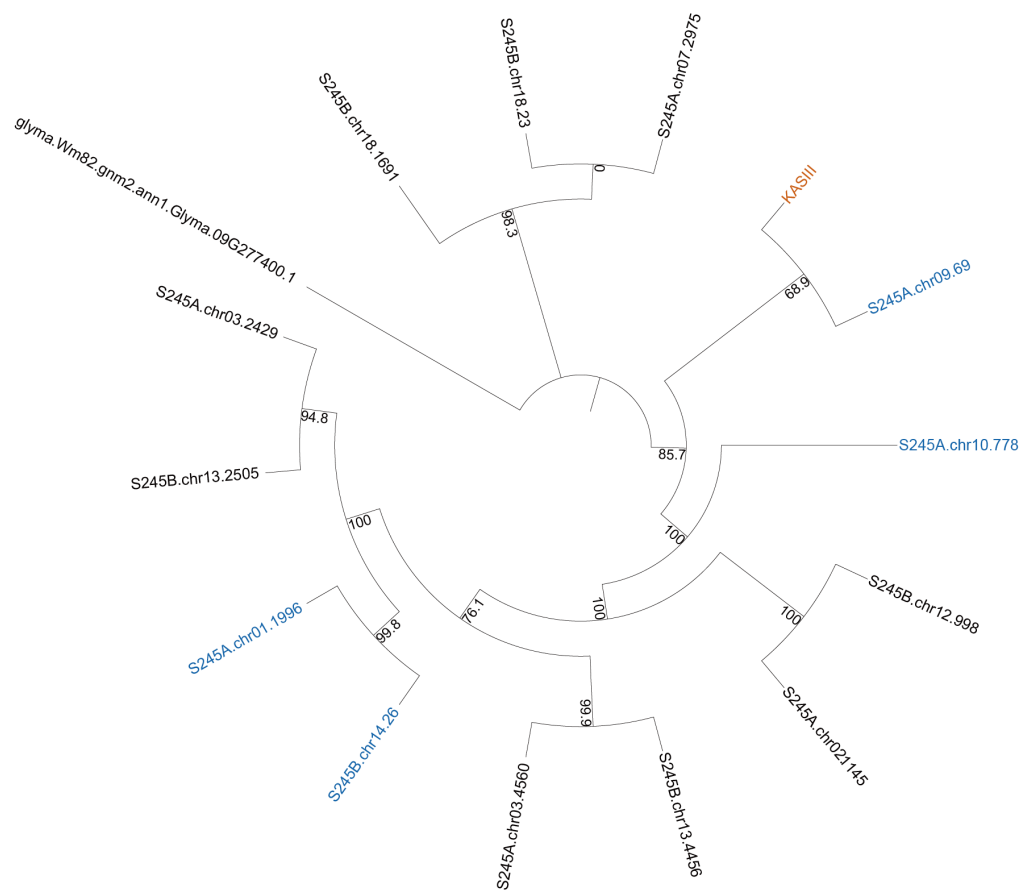


Supplementary Fig. 24 WGCNA analysis and network construction. **a-b**, Evaluation of scale-free fit index at different soft threshold powers. **(a)** Lipid. **(b)** Anthocyanidin. **c-d**, A hierarchical clustering tree displays the coexpression modules identified by WGCNA, together with the assigned merged module colors. **(c)** Lipid. **(d)** Anthocyanidin.

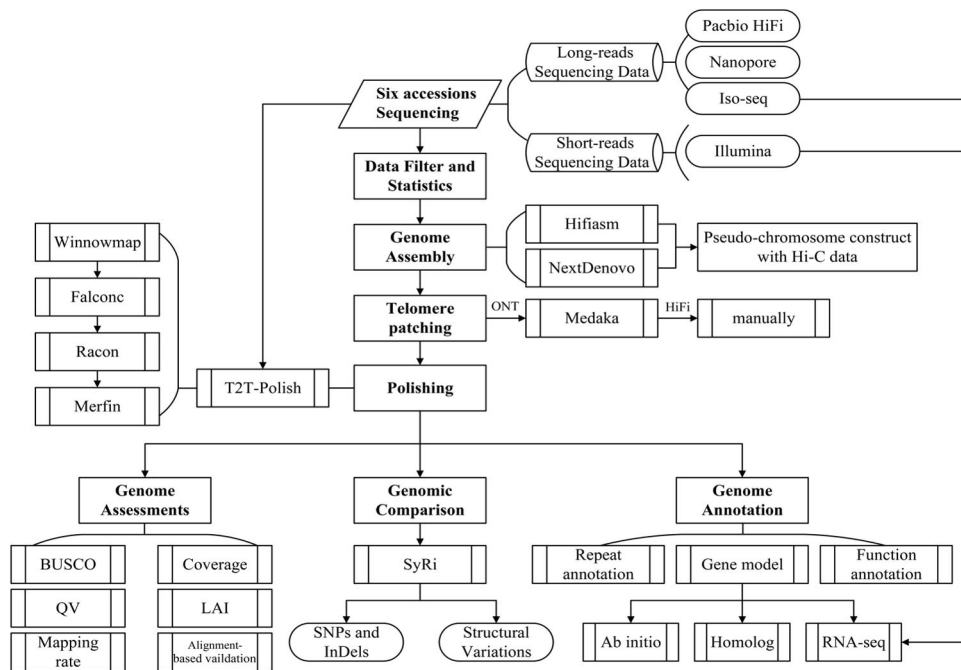
a



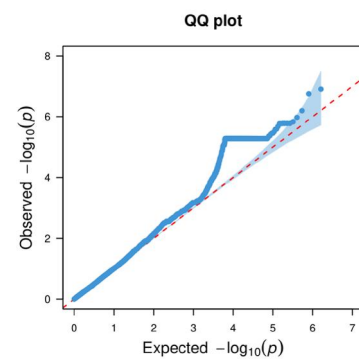
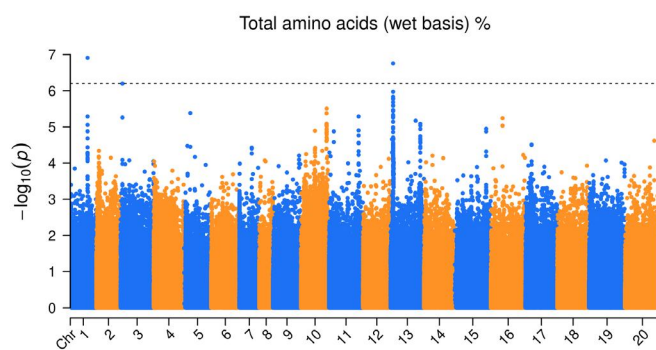
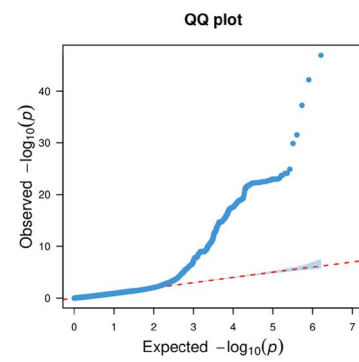
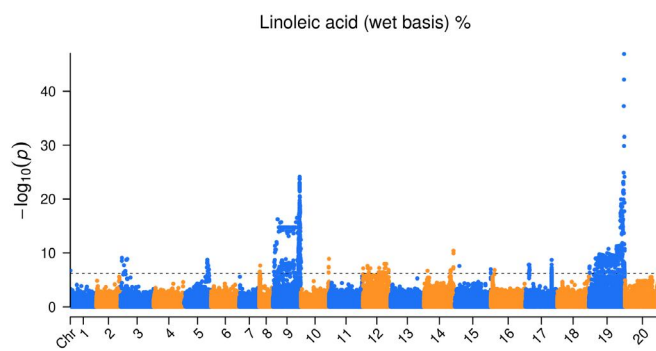
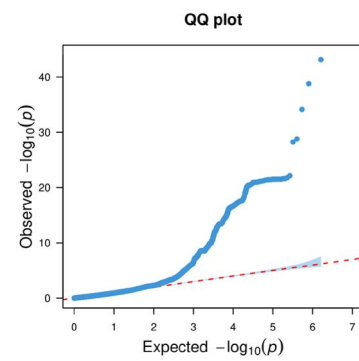
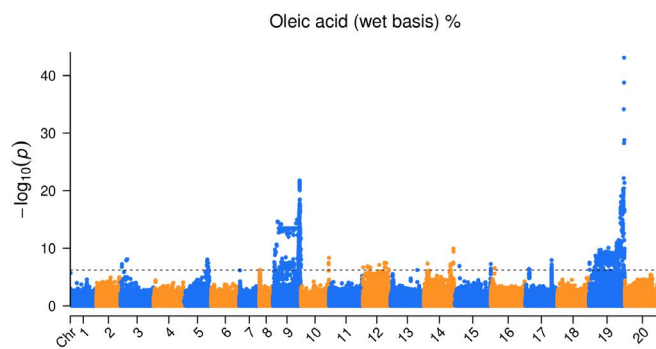
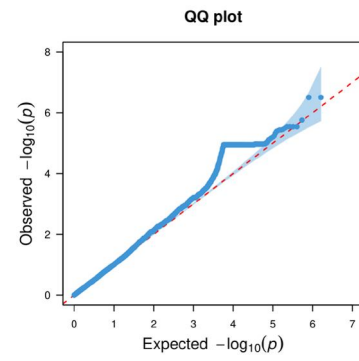
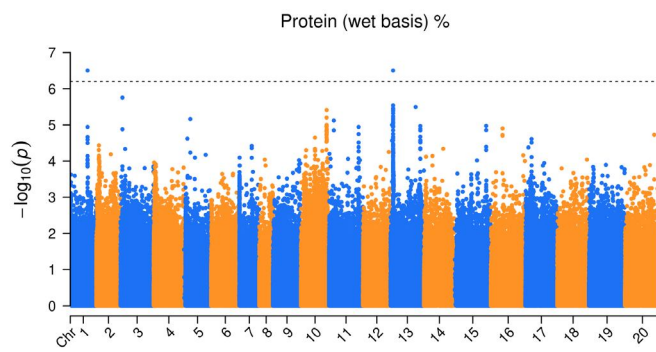
b

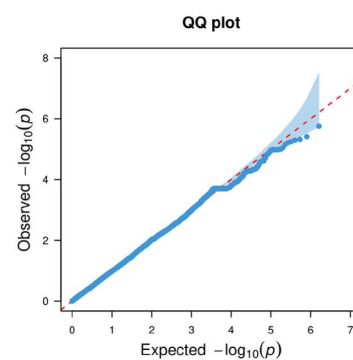
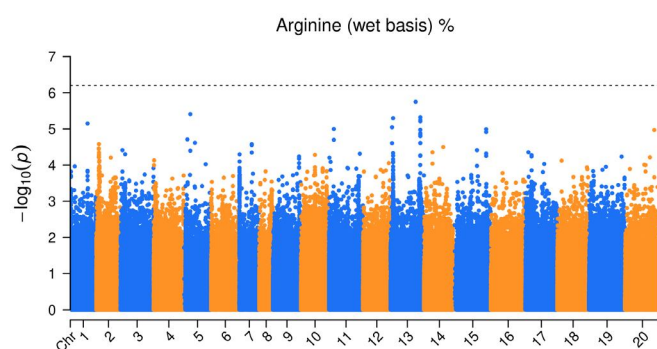
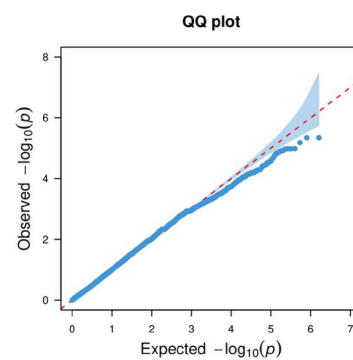
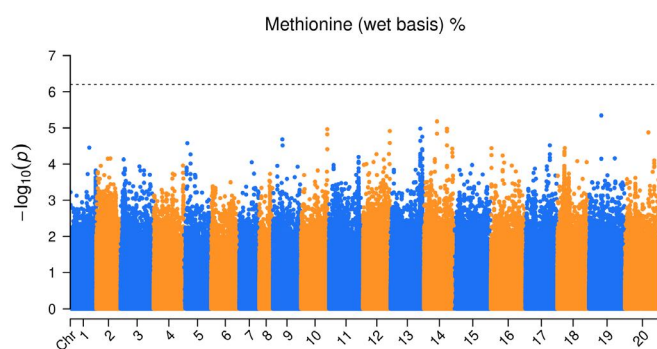
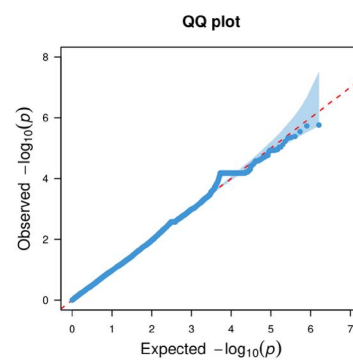
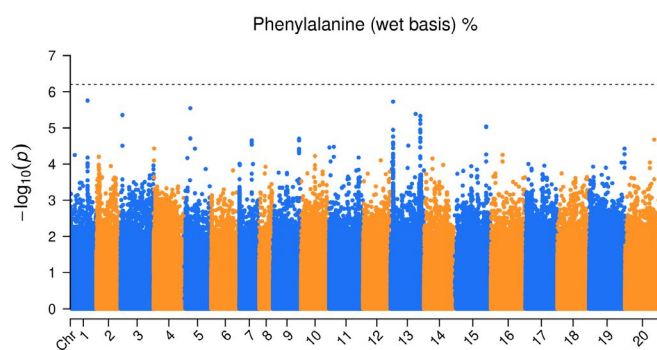
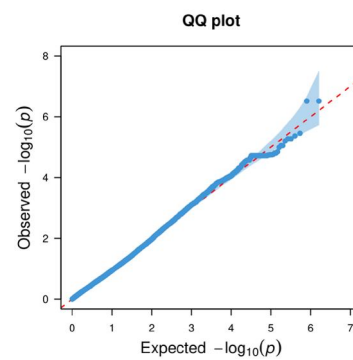
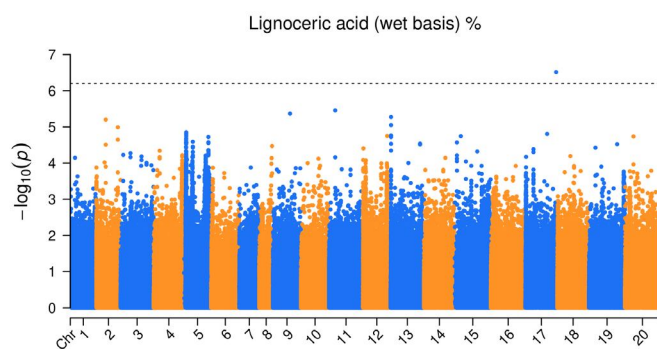


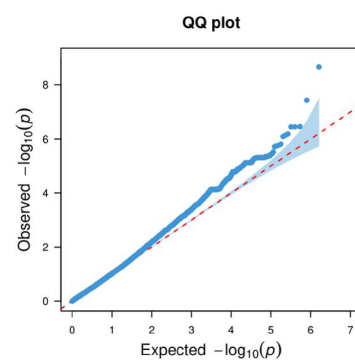
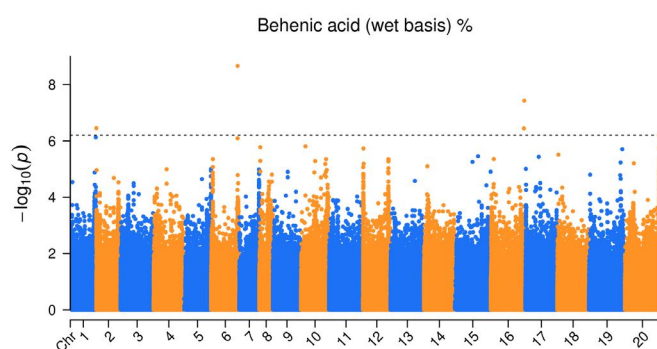
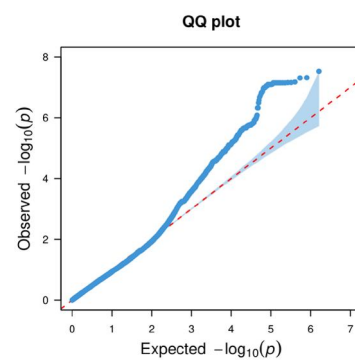
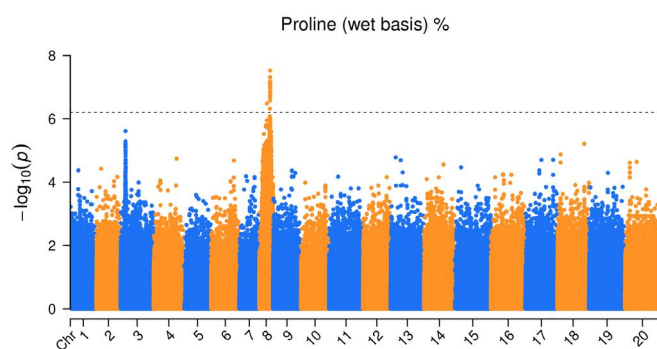
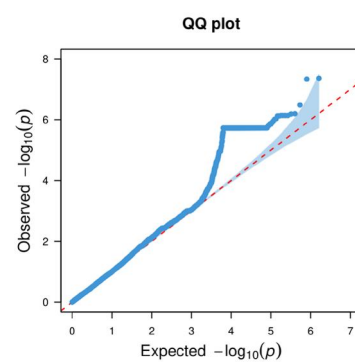
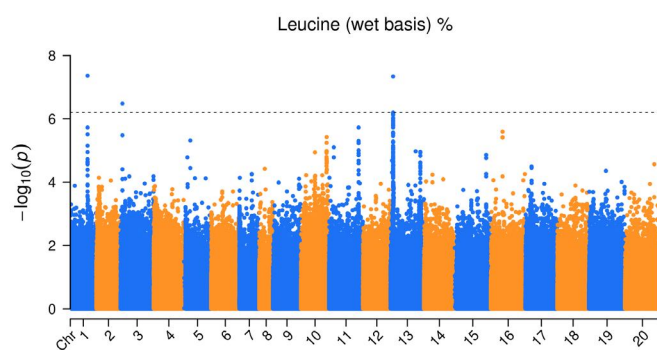
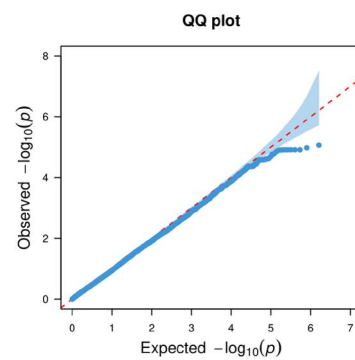
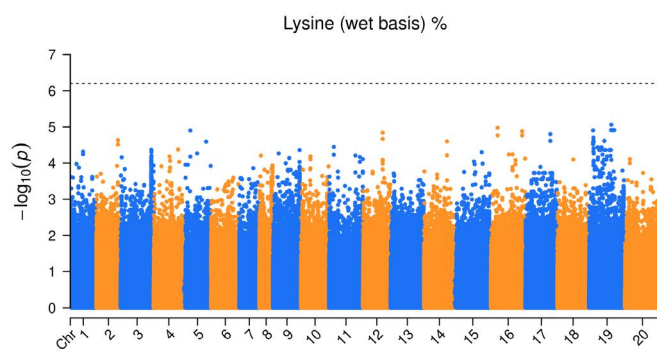
Supplementary Fig. 25 Evolutionary tree of transcription factors and enzymes in the peanut genome. **a**, The tree of FAD genes. **b**, The tree of KAS genes. **c**, The tree of MYB genes. **d**, The tree of bHLH genes. The outgroup of the bHLH tree is the sequence of *Arabidopsis*, and the outgroups of the other trees are soybean. Orange represents the relevant genes that have been verified in peanuts, and blue represents the genes related to anthocyanins and lipids that we screened out.

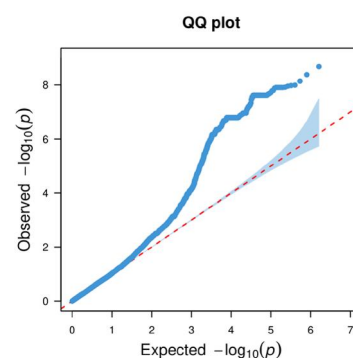
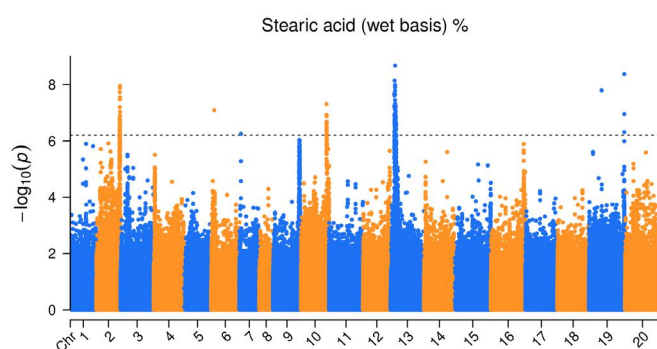
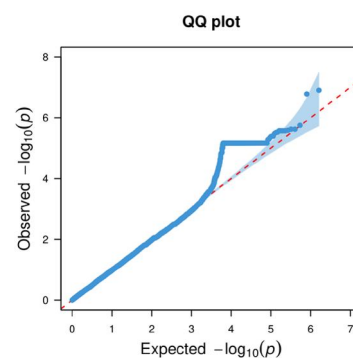
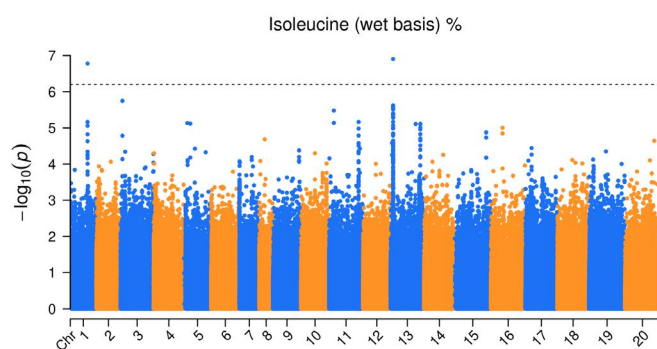
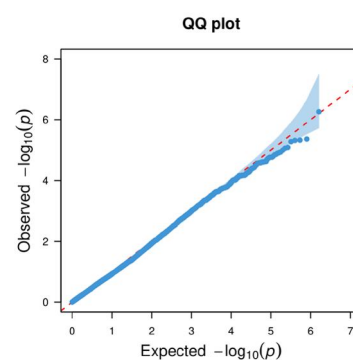
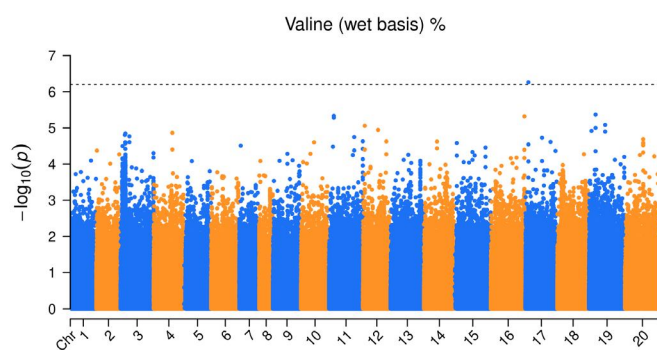
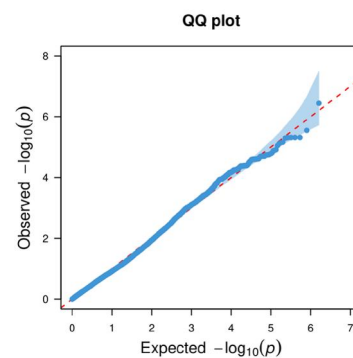
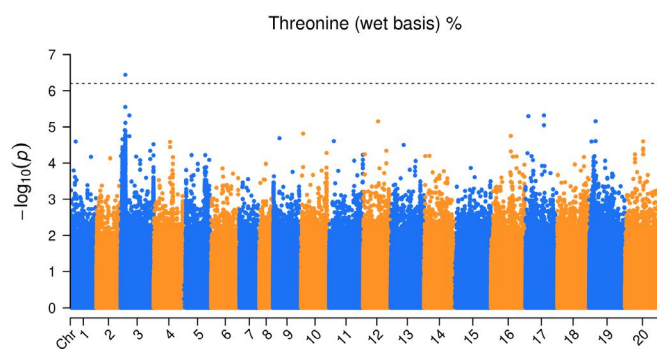


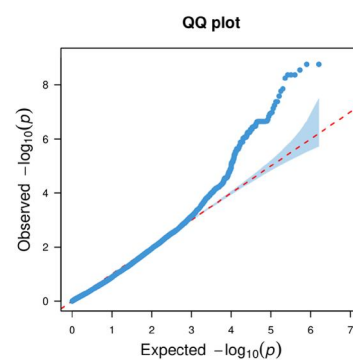
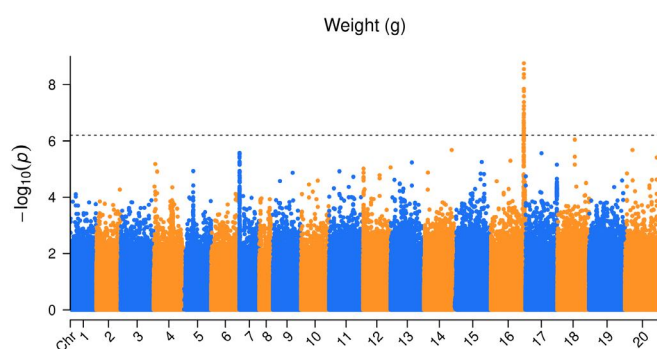
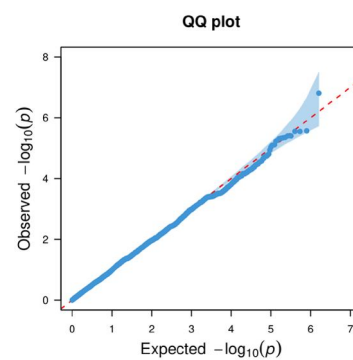
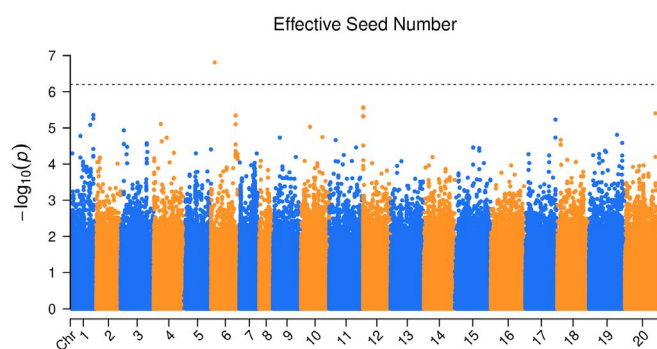
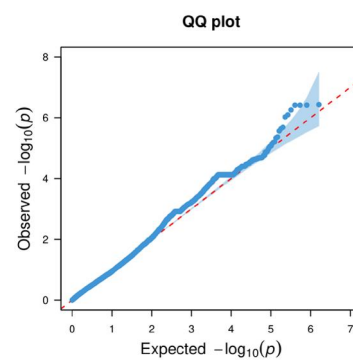
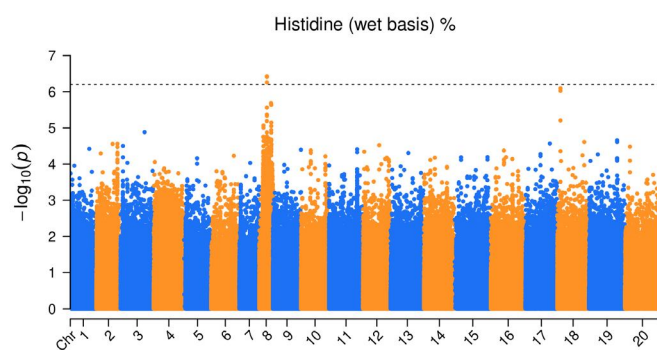
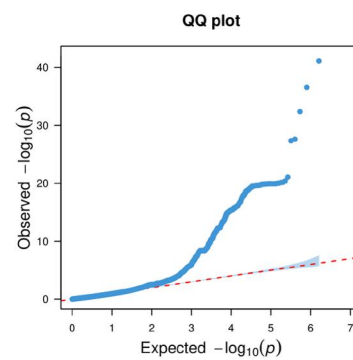
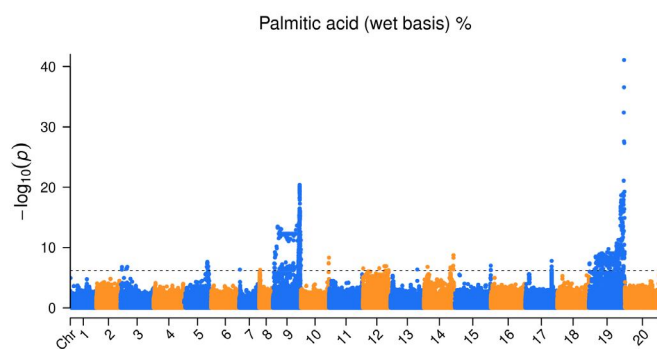
Supplementary Fig. 26 The flowchart of genome assembly.

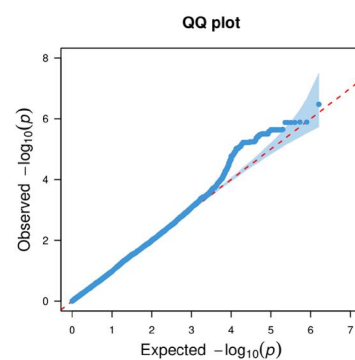
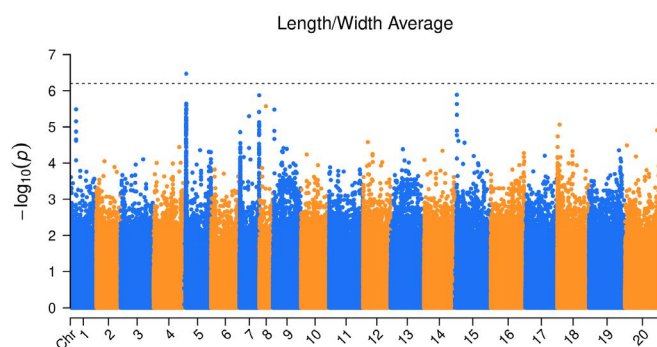
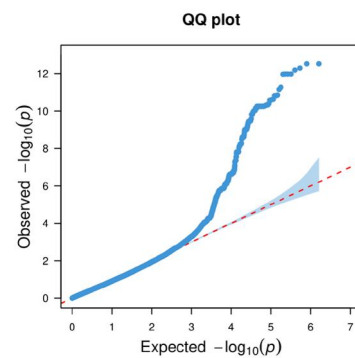
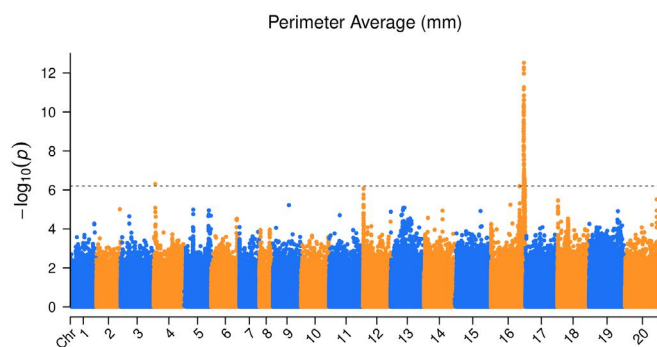
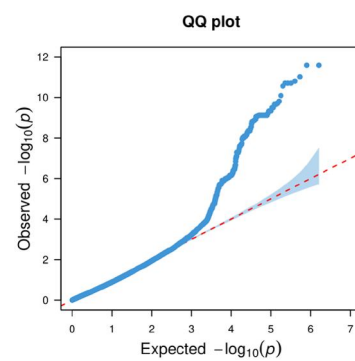
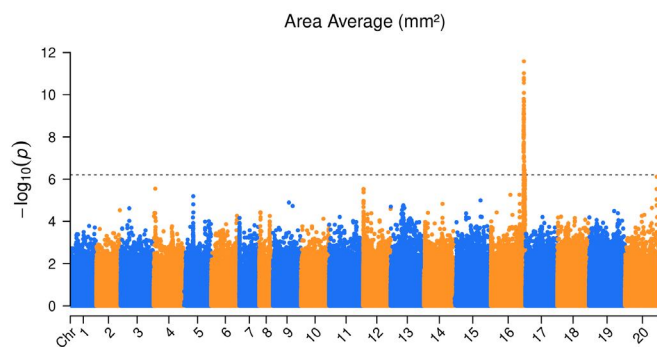
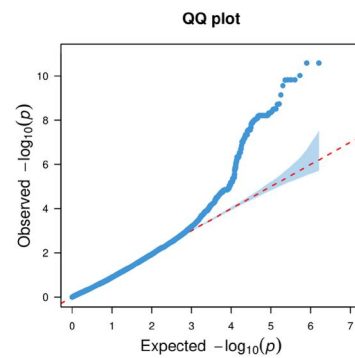
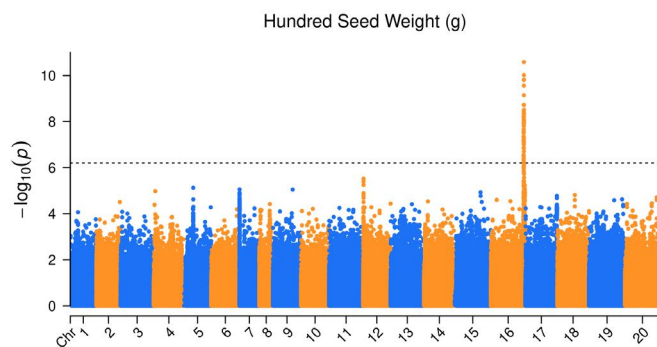


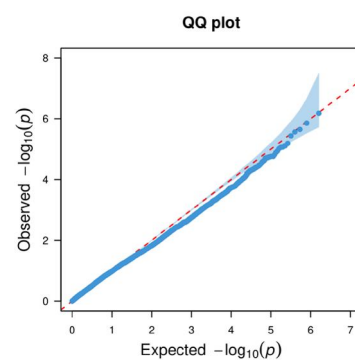
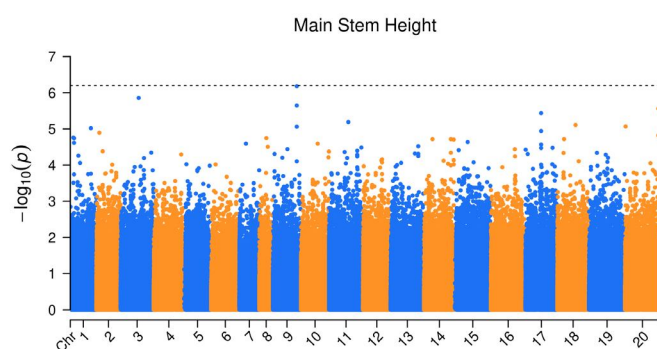
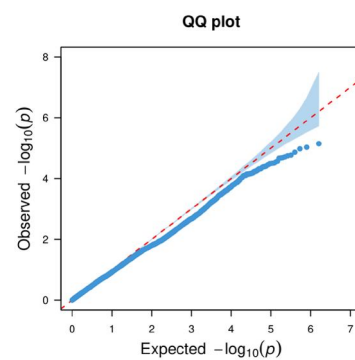
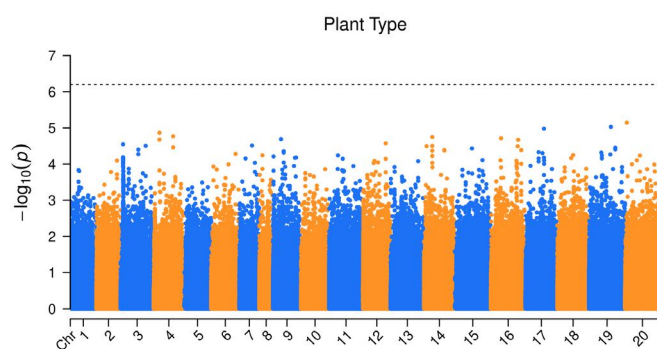
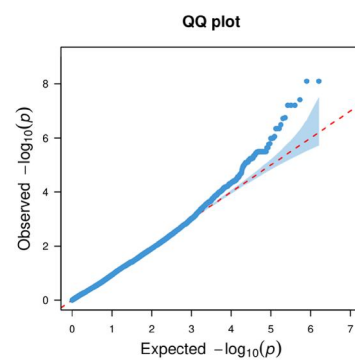
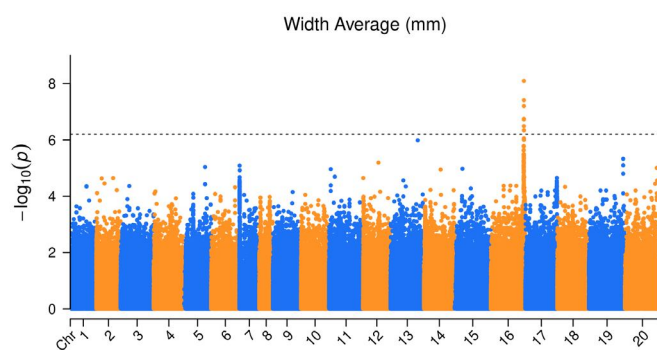
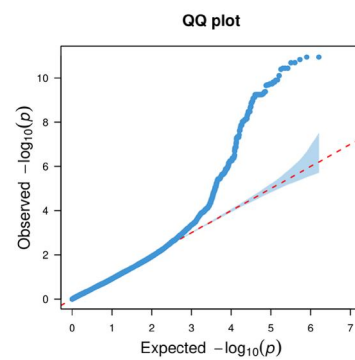
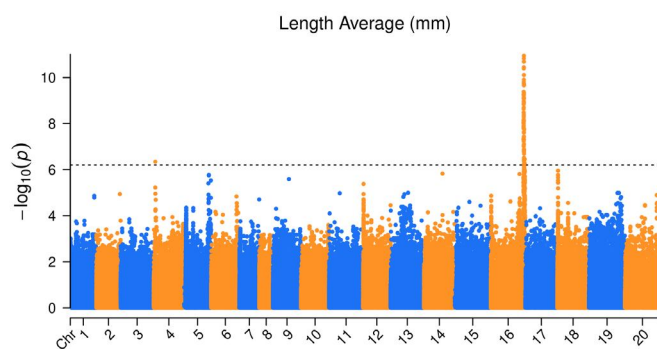


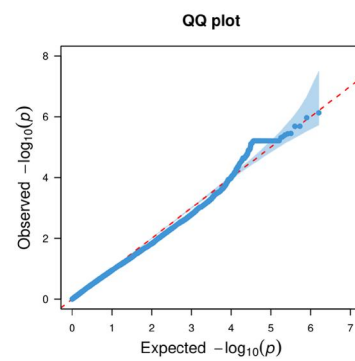
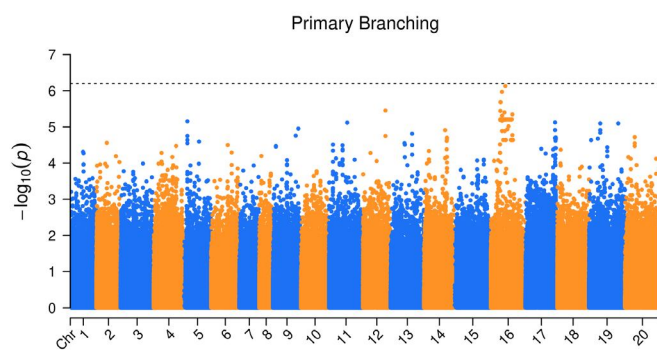
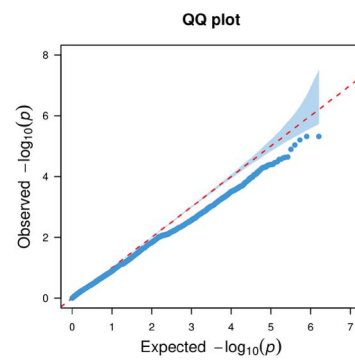
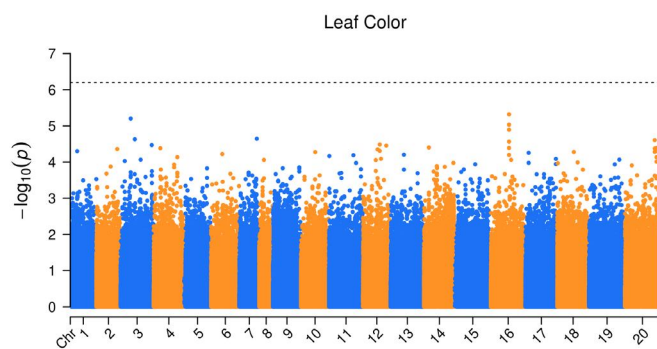
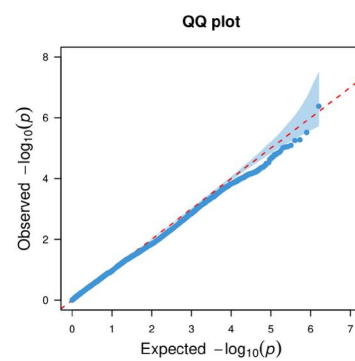
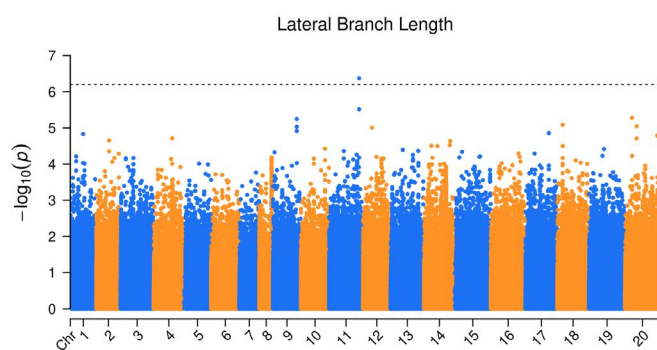
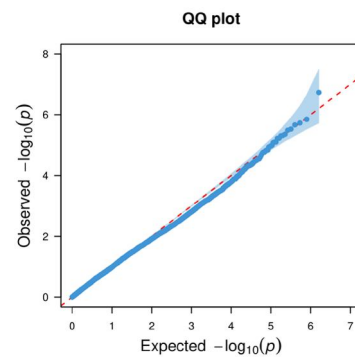
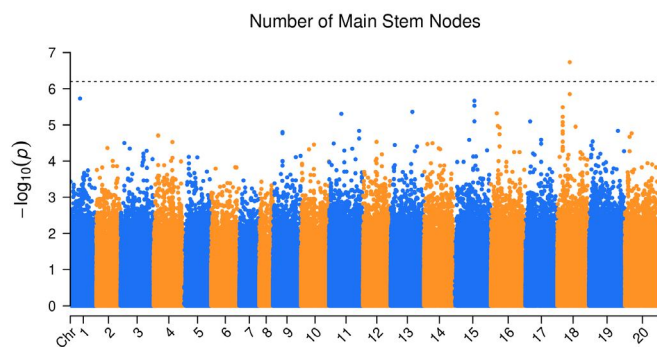


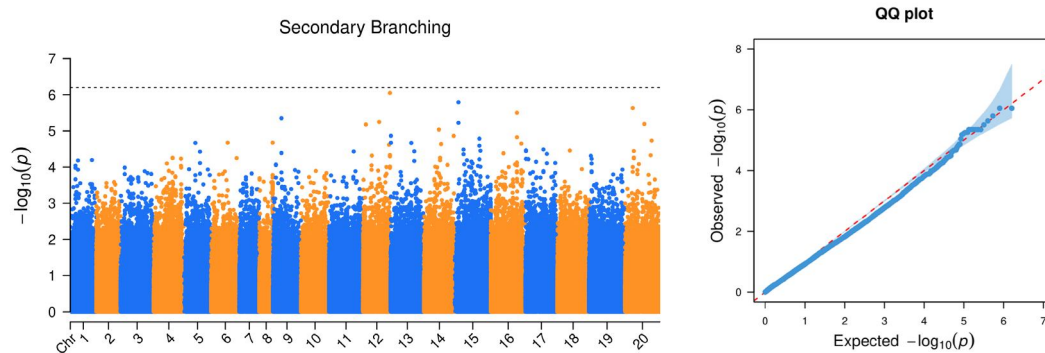












Supplementary Fig. 27 GWAS for 33 traits. In each Fig, the left side depicts the Manhattan plot, while the right side represents the QQ plot. The horizontal dashed line indicates the significant threshold ($-\log_{10}(P)$; $P = 1/\text{SNPs}$).

References

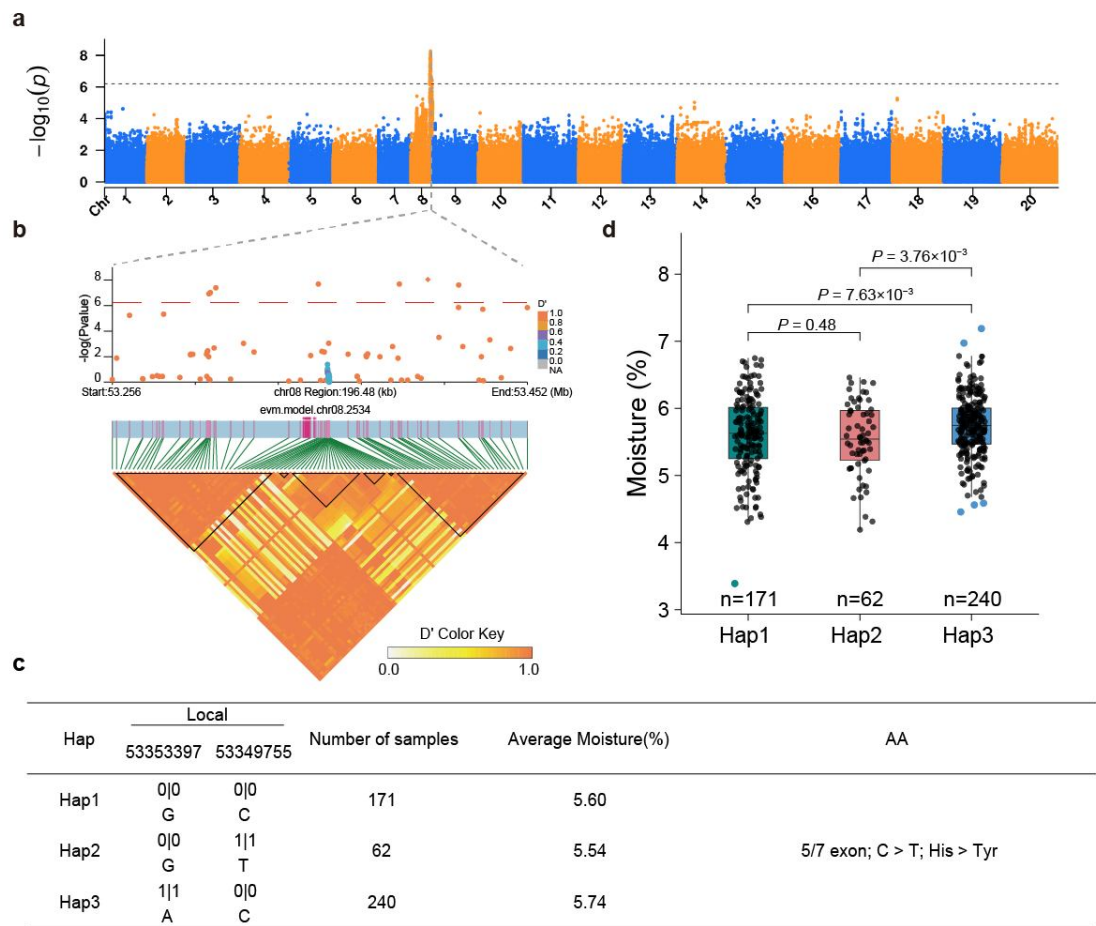
1. Bertoli, D. J. *et al.* The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat Genet.* **51**, 877-884 (2019).
2. Bertoli, D. J. *et al.* The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat Genet.* **48**, 438-446 (2016).
3. Zhao, K. *et al.* Pangenome analysis reveals structural variation associated with seed size and weight traits in peanut. *Nat Genet.* **57**, 1250-1261 (2025).
4. Wang X, *et al.* A telomere-to-telomere genome assembly of the cultivated peanut. *Mol Plant.* **18**, 5-8 (2025).

Supplementary Notes

Candidate gene for Kernel dehydration rate (KDR) in peanut

Due to the subterranean nature of peanut, its seeds tend to have a high moisture content. Therefore, post-harvest, the seeds must be rapidly dried and maintained at a specific moisture level for proper storage, as excessive moisture may lead to germination or mold formation. This is particularly problematic under warm and humid conditions, which may also promote aflatoxin contamination. Therefore, we conducted a genome-wide association study (GWAS) using seed moisture content as the phenotypic trait. A significant signal was identified at the end of chromosome 08 (chr08: 53,353,397) (Supplementary Notes Fig.1a, b). Functional annotation of genes within a 200kb region surrounding this locus revealed two consecutive genes, chr08.2534 and chr08.2535, which were annotated as Ethylene-insensitive proteins (*EIN2*). These genes were designated as *Ah2534_EIN2* and *Ah2535_EIN2*. Previous studies have demonstrated that *EIN2* plays a critical role in ethylene signaling, influencing various developmental processes and responses to multiple stresses, including nutrient deficiency, metal ion stress, plant immunity, and abiotic stressors⁵. Furthermore, Yu et al. (2025) identified a causal micropeptide, microqKDR1, in maize that regulates kernel dehydration by modulating the expression of *EIL1/EIL3*⁶. Knockout experiments involving *ZmEIL1* and *ZmEIL3* showed that KDR levels were significantly reduced in the knockout lines compared to the wild-type, providing further evidence of the involvement of the ethylene signaling pathway in kernel dehydration. Gene structure analysis shows that the peak is located in the promoter region of the *Ah2534_EIN2* gene, 1,073bp upstream of the ATG, where a population-level variation (G to A substitution) was found. Further analysis of the exonic region reveals another population-level variation (C to T substitution) at the fifth exon, which results in an amino acid change from histidine (His) to tyrosine (Tyr). We performed a haplotype analysis of the gene, dividing it into three haplotypes: Hap1 (GC), Hap2 (GT), and Hap3 (AC) (Supplementary Notes Fig.1c). In the population analysis, these three haplotypes contained 171, 62, and 240 accessions, respectively, with corresponding average moisture

contents of 5.60%, 5.54%, and 5.74%. Significance analysis indicated that the moisture content difference between Hap1 and Hap2 was not significant ($P = 0.47$), while Hap3 showed a significantly higher moisture content than both Hap1 ($P = 7.63 \times 10^{-3}$) and Hap2 ($P = 3.76 \times 10^{-3}$) (Supplementary Notes Fig.1d). No population-level variation was found in the gene structure of *Ah2535_EIN2*. Based on these results, we infer that the *Ah2534_EIN2* gene plays a crucial role in regulating the dehydration rate (KDR) and moisture content in peanuts. Among the three haplotypes, Hap2, which exhibited lower moisture content, has significant implications for the subsequent breeding of peanuts with high KDR and low moisture content, which may help reduce the risk of aflatoxin contamination.

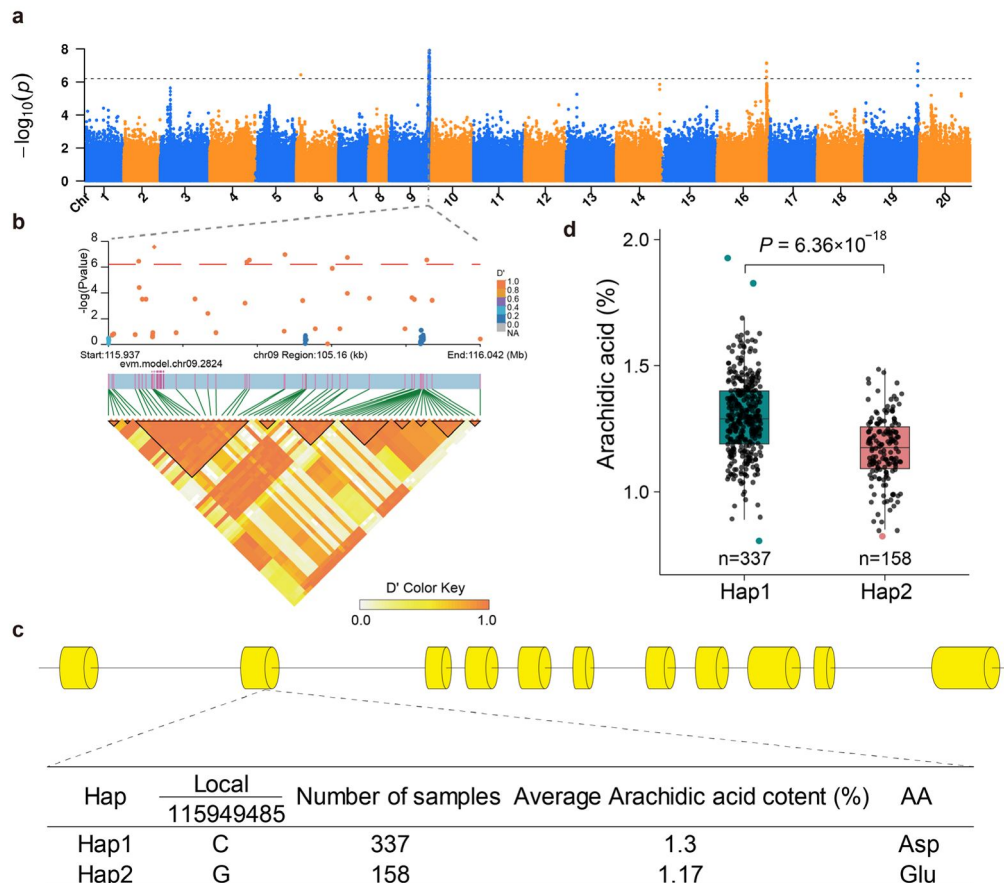


Supplementary Notes Fig. 1 | Candidate gene for kernel dehydration rate. a, Manhattan plot for kernel dehydration rate, significant signals at chr08 ($P = 6.26 \times 10^{-7}$; Bonferroni correction). **b,** LD block of a candidate peak within 200-kb interval regions. **c,** Haplotype analysis of *Ah2534_EIN2* gene. **d,** Comparison of average kernel dehydration

rate differences between Hap1, Hap2, and Hap3 at population levels. Center line, median; box lower and upper edges, 25% and 75% quartiles, respectively; whiskers, $1.5 \times \text{IQR}$; colored dots, outliers. IQR, interquartile range. *P* values were calculated by two-tailed Student's *t*-tests

Candidate gene for arachidic acid content in peanut

Arachidic acid (C20:0), an important component of very long-chain fatty acids (VLCFAs), is associated with the prevalence of atherosclerosis and cardiovascular disease at high levels of VLCFAs⁷. Through GWAS analysis, we identified a significant signal on chromosome 9, with the peak at chr09:115,949,485 (Supplementary Notes Fig.2a, b). Further analysis revealed that a base change from C (Hap1) to G (Hap2) in the second exon of the chr09.2824 gene, annotated as a chaperone protein DNAJ, resulting in an amino acid change (Asp to Glu) (Supplementary Notes Fig.2c). Population analysis showed that 337 samples were of the Hap1 type, with an average arachidic acid content of 1.30%, while 158 samples with the Hap2 type had an average arachidic acid content of 1.17% (Supplementary Notes Fig.2d). It can be speculated that the chr09.2824 gene may have played an important role in regulating the arachidic acid content in peanut seeds.



Supplementary Notes Fig. 2 | Candidate gene for arachidic acid content. **a**, Manhattan plot for arachidic acid content, significant signals at chr09 ($P = 6.26 \times 10^{-7}$; Bonferroni correction). **b**, LD block of a candidate peak within 200-kb interval regions. **c**, Gene structure and haplotype analysis of chr09.2824 gene. **d**, Comparison of average arachidic acid content between Hap1 and Hap2 at population levels. Center line, median; box lower and upper edges, 25% and 75% quartiles, respectively; whiskers, $1.5 \times \text{IQR}$; colored dots, outliers. IQR, interquartile range. P values were calculated by two-tailed Student's t -tests

Significant signals were also identified for other phenotypes (Supplementary Tables 32-34 and Supplementary Fig. 27).

References

- Su, M. & Hou, S. Ethylene insensitive 2 (EIN2) destiny shaper: The post-translational modification. *J Plant Physiol.* **295**, 154190 (2024).
- Yu, Y. et al. A Zea genus-specific micropeptide controls kernel dehydration in maize. *Cell.*

188, 44-59 (2025).

7. Huai, D. et al. Enhancing peanut nutritional quality by editing AhKCS genes lacking natural variation. *Plant Biotechnol J.* **22**, 3015–3017 (2024).