

# Pan-genome-based resequencing of 2,320 accessions reveals structural variations and accelerates breeding advances in cultivated peanut

Received: 14 July 2025

Accepted: 19 August 2026

Published online: 15 September 2026

 Check for updates

Yiyang Liu<sup>1,8</sup>, Weitao Li<sup>1,8</sup>, Rongchong Li<sup>1,8</sup>, Manish K. Pandey<sup>2,8</sup>, Bo Bai<sup>1</sup>, Yan Han<sup>1</sup>, Guiying Tang<sup>1</sup>, Lei Zhang<sup>1</sup>, Annapurna Chitkineni<sup>3</sup>, Yanjiao Li<sup>1</sup>, Libing Li<sup>1</sup>, Shulong Li<sup>1</sup>, Vanika Garg<sup>3</sup>, Fengping Du<sup>1</sup>, Feng Cui<sup>1</sup>, Liangqiong He<sup>4</sup>, Lei Shan<sup>1</sup>, Fangji Xu<sup>1</sup>, Pingli Xu<sup>1</sup>, Ronghua Tang<sup>4</sup>, Reyazul R. Mir<sup>3</sup>, Feng Guo<sup>1</sup>, Xinguo Li<sup>1</sup>, Jialei Zhang<sup>1</sup>, Zheng Zhang<sup>1</sup>, Kuldeep Singh<sup>5</sup>, Qiang He<sup>6</sup>, Guowei Li<sup>1</sup>✉, Xinyou Zhang<sup>7</sup>✉, Rajeev K. Varshney<sup>3</sup>✉ & Shubo Wan<sup>1</sup>✉

The cultivated peanut is a crucial global legume crop that is essential for food security and nutrition, particularly in developing regions. However, its limited genetic variation hampers breeding progress and yield improvement. Here we constructed a graph-based pan-genome for peanut, incorporating 14 genomes that represent all 6 peanut varieties. Using this pan-genome, we genotyped 2,320 accessions, covering 88.03% of ICRISAT and 59.21% of USDA core germplasm, enriching valuable resources for genomic studies and breeding. We cataloged genomic structural variations and investigated the role of homoeologous exchanges in population divergence. Through our pan-genome approach, we overcame the challenges of genotyping posed by homoeologous exchanges and identified key genes associated with flowering and dwarfism in peanut. By integrating superior haplotypes and germplasm resources guided by the pan-genome, we further developed high-yield dwarf lines. This work provides essential genomic resources to accelerate functional gene discovery and modern peanut breeding.

Cultivated peanut is an allotetraploid species (AABB) derived from a single hybridization and polyploidization event involving two diploid progenitors, *Arachis duranensis* and *Arachis ipaensis*<sup>1,2</sup>. Based on phenotypic variation, cultivated peanut is classified into two subspecies and six botanical varieties: subsp. *hypogaea*, comprising var. *hypogaea* and var. *hirsuta*; and subsp. *fastigiata*, comprising var. *fastigiata*, var. *vulgaris*, var. *peruviana* and var. *aequatoriana*<sup>3</sup>. Cultivated peanut is believed to have been domesticated from its wild tetraploid progenitor in the region of southern Bolivia and

northern Argentina, with subsequent spread across the globe through migration routes<sup>4</sup>. However, the cultivated peanut has experienced substantial genetic bottlenecks throughout its evolution and domestication, resulting in narrow genetic diversity<sup>5,6</sup>. Exploitation of genetic diversity is critical for genetic improvement of peanut and modern breeding to adapt to growth in production demand<sup>7,8</sup>.

In this study, we constructed a peanut pan-genome using 14 genomes representing all 6 botanical varieties. This comprehensive resource enabled systematic identification of a broad spectrum

A full list of affiliations appears at the end of the paper. ✉e-mail: [ligw\\_saas@163.com](mailto:ligw_saas@163.com); [haasz@126.com](mailto:haasz@126.com); [rajeev.varshney@murdoch.edu.au](mailto:rajeev.varshney@murdoch.edu.au); [wanshubo2016@163.com](mailto:wanshubo2016@163.com)

of structural variations (SVs) and assessment of their contributions to agronomically important traits. By integrating whole-genome sequencing data from 2,320 geographically and genetically diverse germplasm accessions, we generated the most comprehensive peanut pan-genomic resource to date. This resource enabled comprehensive SV discovery and improved gene mapping in homoeologous exchange (HE) regions; it also allowed us to identify loci underlying key agronomic traits, including flowering time and dwarfism, providing new insights into the genetic basis of peanut domestication and improvement. Through pan-genome-guided integration of superior haplotypes, we developed a superior dwarf line with high-yield potential. This study provides a comprehensive biodigital resource to accelerate the development of high-yield peanut cultivars.

## Results

### Genome assembly and pan-genome analysis of cultivated peanut

To represent the genome diversity of cultivated peanut, we selected 10 representative accessions of all 6 botanical varieties from the phylogenetic relationships of 401 accessions (Fig. 1a, Supplementary Figs. 1 and 2, Supplementary Table 1 and Supplementary Note 1). We performed de novo genome assembly of the 10 accessions using high-fidelity (HiFi) reads (average coverage, 45.20×; Supplementary Table 2), followed by high-throughput chromatin conformation capture (Hi-C) scaffolding using paired-end reads (average coverage, 110.61×; Supplementary Table 3) to correct misassemblies and anchor contigs to 20 pseudochromosomes. For genome polishing, we used short-read sequencing data with an average depth of 105.70× (Supplementary Table 4). Five genomes were further improved with respect to integrity and continuity, with an average of 63.12 Gb of ultralong Oxford Nanopore Technologies (ONT) data (Supplementary Table 5). We generated 10 reference-quality genomes with an average assembly size of 2.61 Gb, contig N50s of 70.87–140.38 Mb, and scaffold N50s of 132.29–141.62 Mb (Supplementary Table 6). HiFi and short-read mapping rates averaged 99.82% and 99.93%, respectively, and BUSCO completeness averaged 99.27% (Supplementary Tables 7–10). Furthermore, *K*-mer completeness averaged 98.70%, with a quality value of 51.81, and the long terminal repeat (LTR) assembly index exceeded 10 for all assemblies, supporting their reference quality (Supplementary Tables 10 and 11). We annotated an average of 78,222 protein-coding genes per genome, with 99.28% BUSCO completeness, using *ab initio*, homology-based and RNA sequencing (RNA-seq)-based predictions (Supplementary Table 10). Repetitive sequences comprised 78.36% (–2.0 Gb) of each genome, predominantly LTR retrotransposons, with Gypsy elements accounting for 41.85% (Supplementary Table 12).

We conducted a comprehensive pan-genome analysis for cultivated peanut species with the ten newly assembled genomes and four previously published genomes<sup>2,9–11</sup>. Before constructing the pan-genome gene set, we employed a cross-mapping approach to standardize gene structure annotations and thus enhance the consistency of gene clustering. Following reannotation, the average number of genes across the 14 genomes was 82,543 (Supplementary Table 13). An orthologous gene analysis subsequently classified all 1,155,597 genes from the 14 genomes into 88,733 gene families, identifying 17,591 additional families compared with a recent pan-genome study of eight peanut germplasms<sup>12</sup>. Among the gene families, 48,948 were defined as core gene families present in all 14 accessions, 6,427 were softcore gene families present in 13 accessions, 31,541 were dispensable gene families present in 2 to 12 accessions and 1,817 were private gene families found in only 1 accession (Fig. 1b). In addition, there were 11,789 genes (1.02% of the total genes) that were not assigned to any orthogroup. Although dispensable and private gene families represented a substantial portion (37.60%) of all gene families identified across the 14 accessions, the genes belonging to these families accounted for an average of only

19.70% of the total genes within each individual accession (Fig. 1c and Extended Data Fig. 1a). Compared with dispensable genes, core and softcore genes had markedly lower nonsynonymous/synonymous substitution ratios ( $K_a/K_s$ ) and were expressed at significantly higher levels (Extended Data Fig. 1b,c). Furthermore, Kyoto Encyclopedia of Genes and Genomes enrichment analysis revealed that core genes were predominantly associated with cell signaling and metabolism regulation pathways, whereas dispensable genes were mainly involved in pathways such as DNA repair (Extended Data Fig. 1d). It is typically suggested that core genes, conserved during evolution, are needed to maintain fundamental biological functions, whereas dispensable genes may have undergone more rapid evolution as part of adaptation to varying conditions.

### SV characterization and graph-based genome

The reference genomes allowed us to identify SVs by comparative genomic analysis among different accessions. Alignment of the 13 genomes to the Tifrunner genome identified a total of 126,758 SVs, including 119,965 presence/absence variations (PAVs), 5,639 copy number variations (CNVs), 614 inversion events and 540 translocation events (Fig. 1d, Supplementary Fig. 3 and Supplementary Table 14). Each accession contained approximately 34.00 Mb of inversion events and 1–8 intrachromosomal translocations (size range: 0.20–10.11 Mb) across chromosomes (Supplementary Tables 15 and 16). Specifically, based on the phylogeny of the 14 accessions, we identified an inversion of approximately 5.76 Mb on chr. 13 that effectively distinguished the two subspecies of cultivated peanut (Fig. 1e and Supplementary Fig. 4). We randomly selected 110 SVs (ranging from 628 bp to 11,452,067 bp) and validated their authenticity using HiFi, Hi-C or PCR amplification (Supplementary Figs. 5–8 and Supplementary Table 17).

We subsequently merged 126,758 SVs into a set of 44,073 non-redundant SVs and constructed a graph-based pan-genome by integrating the identified SVs into the Tifrunner v.2 genome sequences. PAVs accounted for approximately 90% of these SVs, and about 30% of them were located in genic regions (within coding sequences or introns) and flanking regions (upstream or downstream of genes) (Extended Data Fig. 2a,b). To assess the impact of these PAVs on gene expression, we performed a differential expression analysis between genes with and without PAVs in the corresponding genomic regions across each accession. We found that PAVs located in the coding regions were more likely to affect gene expression than those located in introns, upstream and downstream of genes (Extended Data Fig. 2c). Furthermore, transposable elements (TEs) were enriched in PAVs, with LTR-TE-derived PAVs comprising the largest proportion (49.1%) of all PAVs (Extended Data Fig. 2d,e). Although TE-derived PAVs were more abundant in intergenic regions, their distribution within intragenic regions was similar to that of non-TE-derived PAVs (Extended Data Fig. 2f).

### Population structure and diversity based on pan-genome

To explore genetic diversity at the population scale based on the pan-genome, we resequenced 2,320 cultivated peanut accessions with an average coverage of 14× (Supplementary Table 18). These accessions included 782 var. *hypogaea*, 127 var. *hirsuta*, 553 var. *fastigiata*, 792 var. *vulgaris*, 63 var. *peruviana* and 3 var. *aequatoriana* collected from 87 countries or regions (Fig. 2a and Supplementary Table 19). After aligning the resequencing data to the graph-based pan-genome, we identified 29,195,792 single nucleotide polymorphisms (SNPs) and insertions/deletions (indels). Then, we performed phylogenetic analysis and principal component analysis using 16,318,617 quality-filtered SNPs and classified the 2,320 accessions into 4 groups (Fig. 2b,c and Supplementary Table 20). Moreover, structural analysis of 12,816 four-fold degenerate (4DTV) sites revealed a similar clustering pattern (Extended Data Fig. 3a). The four groups were as follows: the PA group, consisting primarily of var. *peruviana* and var. *aequatoriana* accessions; the

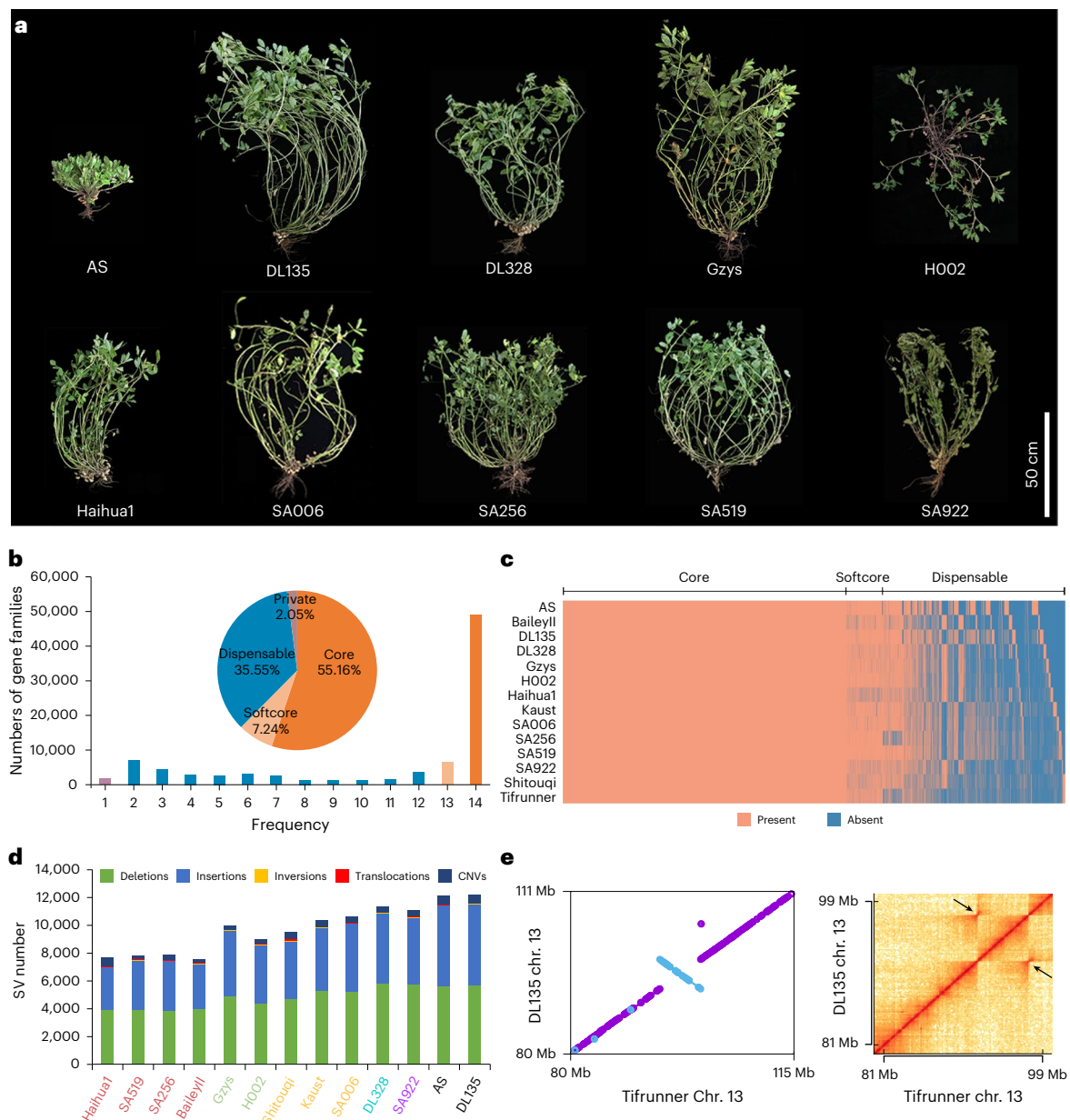


**Fig. 1 | Pan-genome analysis of 14 genomes. a**, Highly diverse agronomic phenotypes among ten peanut accessions, including plant height and architecture. **b**, Compositions of the pan-genome. The histogram shows the frequency of gene families, and the pie chart depicts proportions of core, softcore, dispensable and private gene families. **c**, Presence and absence

information of pan-gene families in the 14 peanut genomes. **d**, Numbers of each SV type (deletions, insertions, inversions, translocations and CNVs) in individual accessions relative to the Tifrunner reference genome. **e**, Collinearity- and Hi-C-validated 5.76-Mb inversion event on chr. 13 between subsp. *hypogaea* (Tifrunner) and subsp. *fastigiata* (DL135).

HH group, which included most of the var. *hypogaea* and var. *hirsuta* accessions; and the Vu and Fa groups, containing most of the var. *vulgaris* and var. *fastigiata* accessions, respectively. Among these four groups, the PA group presented the greatest genetic divergence (fixation index;  $F_{ST}$ ) from the other three groups in pairwise comparisons, as well as the highest nucleotide diversity, with  $\pi = 1.23 \times 10^{-4}$  (Extended Data Fig. 3b).

TreeMix and  $D$ -statistics analyses detected directional gene flow from Fa into PA, as well as from PA and HH into Vu (Extended Data Fig. 3c,d), suggesting that peanut diversification involved postdivergence genetic exchange in addition to lineage divergence. Genome-wide  $f_d$  scans further resolved these signals at the chromosomal level (Supplementary Fig. 9 and Supplementary Table 21). In the Fa-to-PA comparison, two prominent  $f_d$  peaks were detected on chromosome A05 (Extended Data Fig. 3e), overlapping with previously

reported major quantitative trait loci (QTLs) controlling formation of multiseeded pods (3–4 seeds per pod)<sup>13</sup>.

Given the pronounced morphotype divergence between subsp. *hypogaea* and subsp. *fastigiata*<sup>3,14</sup>, we performed a case-control genome-wide association study (GWAS) to identify PAVs associated with morphotype divergence between the two subspecies. This analysis identified 1,967 significant PAVs within the top 5% of GWAS signals (Supplementary Table 22). We also calculated  $F_{ST}$  values between the two subspecies to identify genomic regions associated with subspecies divergence. Genomic windows within the top 5% of  $F_{ST}$  values (covering 147.20 Mb in total) were defined as highly differentiated regions (Supplementary Table 23). By taking the intersection of PAV coordinates with these regions, we identified 5,381 subspecies-differentiated PAVs located within highly diverged genomic regions. Finally, by taking

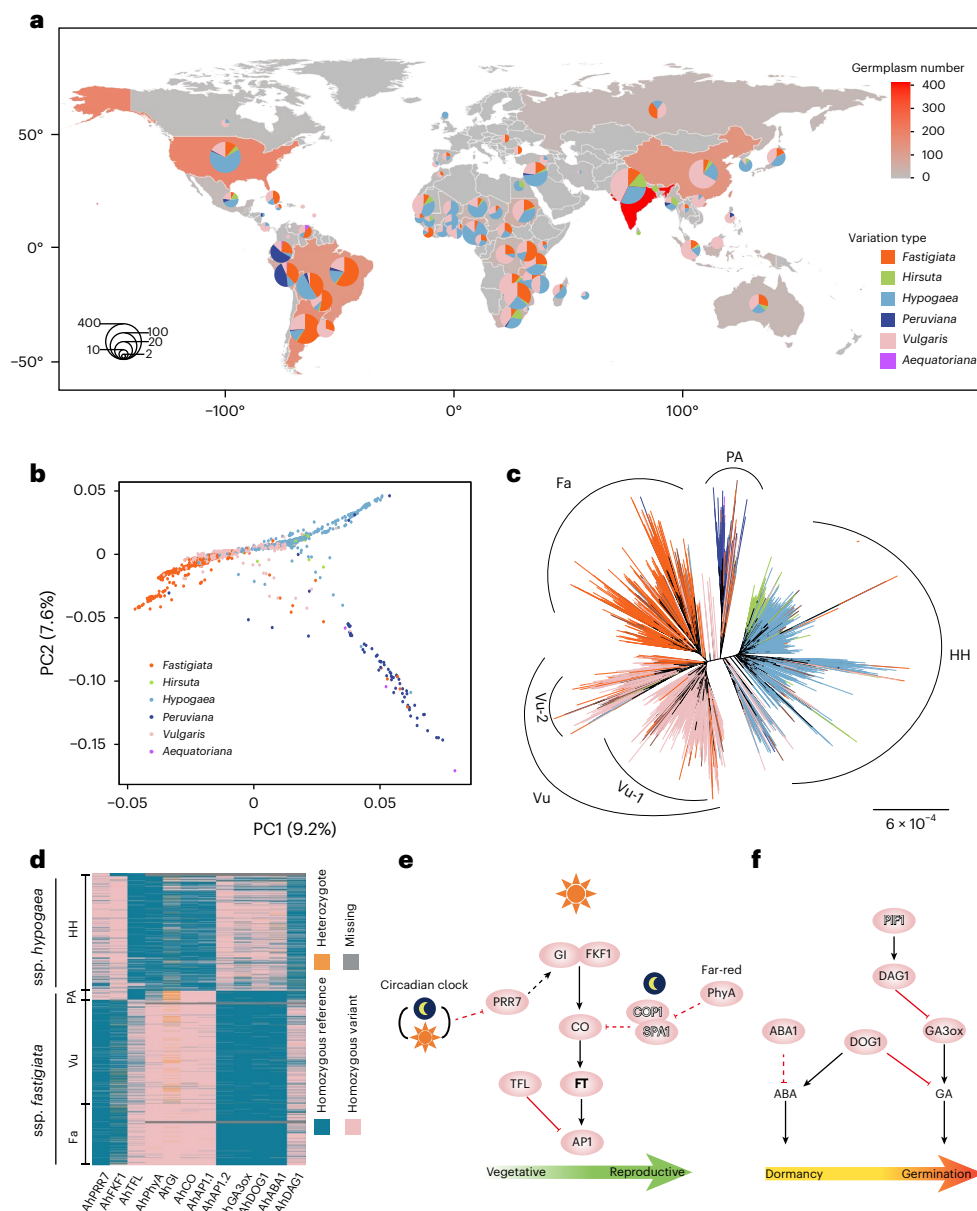


**Fig. 2 | Geographic distribution and phylogenetic analysis of 2,320 resequencing peanut accessions. a**, Geographic distribution of the 2,320 accessions. **b**, Principal component analysis of 2,320 peanut accessions based on SNP information. **c**, Maximum-likelihood phylogenetic tree of peanut accessions. Branches are color-coded based on the variation types. Major clusters are labeled as Fa (*var. fastigiata*), PA (*var. peruviana* and *var. aequatoriana*),

HH (*var. hypogaea* and *var. hirsuta*) and Vu (*var. vulgaris*). **d**, PAVs of 12 genes among 2,320 accessions of different subspecies and variety. **e, f**, Schematic representation of the regulatory network controlling flowering (**e**) and dormancy (**f**) in peanut. ABA, abscisic acid. The base map in **a** was plotted using the R package maps (v.3.4.3; <https://github.com/adeckmyn/maps>), with data derived from Natural Earth (<http://www.naturalearthdata.com>).

the overlap of these with the case-control GWAS signals, we obtained 312 high-confidence PAVs that were both statistically associated and genomically differentiated between subspecies. Among these, 133 PAVs overlapping 157 genes were located within genic and gene upstream/downstream regions, potentially leading to specific gene alterations (Supplementary Table 24). Functional annotation and pathway mapping revealed that several of these genes were involved in pathways related to photoperiod, flowering and seed dormancy (Fig. 2d–f and Supplementary Table 25). For example, *Ah09g143300*, a homolog of the soybean *PhyA* genes (*PHYTOCHROME A*) that function in photoperiod<sup>15</sup>, and *Ah08g346000*, a homolog of *Arabidopsis* *DAG1* (*DOF AFFECTING GERMINATION1*) associated with seed dormancy<sup>16</sup>, both harbored deletions in their coding regions specifically in *ssp. fastigiata* accessions (Extended Data Fig. 3f).

## HE in peanut population

Previous studies have suggested that cultivated peanuts, as segmental allotetraploids, undergo HE with distinct patterns across accessions, and that this contributes to phenotypic variation<sup>1,2,17</sup>. However, the specific role of HE in peanut evolution has remained largely unexplored. Therefore, we investigated the contribution of HE to subspecies divergence by calculating HE values from 586 accessions with sequencing depths  $\geq 20\times$ . All 586 accessions exhibited widespread HE events, as evidenced by our identification of 3,352 HE genes among 17,373 gene pairs (Supplementary Table 26). HEs were strongly biased toward B-to-A exchanges, in which B chromosome segments replaced A chromosome segments, with more events (95 versus 13) and a greater cumulative exchange length (33.21 Mb versus 11.26 Mb) than A-to-B exchanges (Fig. 3a,b and Supplementary Table 27). This pattern was observed

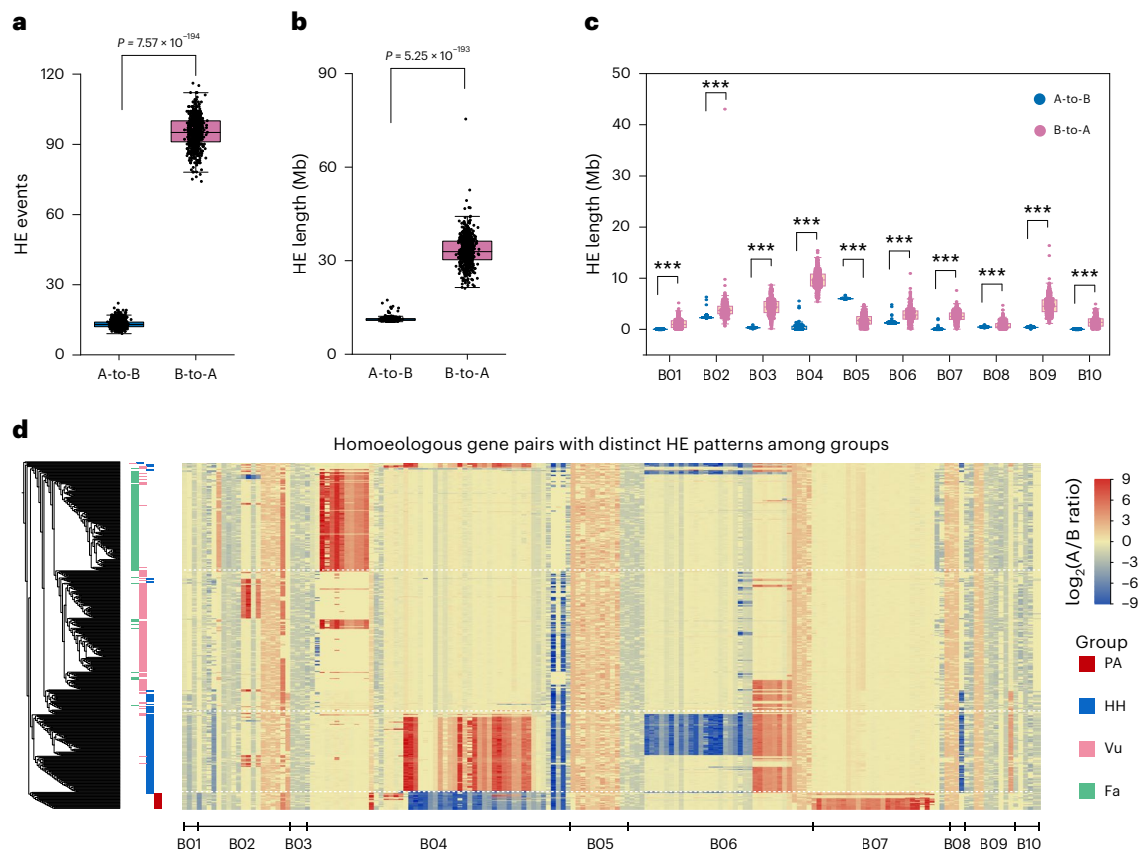


**Fig. 3 | HE variation in peanut. a**, Comparison of A-to-B and B-to-A HE events. **b**, Comparison of HE length between A-to-B and B-to-A events. **c**, A-to-B and B-to-A HE length across the B genome corresponding to chr. 11 to chr. 20 of the cultivated peanut. In **a–c**, data are from 586 accessions ( $n = 586$  accessions). Box plots show the median (center line), and the first and third quartiles (lower and upper box edges, respectively), with whiskers extending to 1.5 times

the interquartile range. **d**, Clustering tree generated using average linkage hierarchical clustering based on Euclidean distance calculated from the HE values of 3,352 gene pairs among the 586 accessions. PA, HH, Vu and Fa represent the four groups clustered based on SNPs. The heatmap showed distinct HE patterns among groups. Statistical significance was assessed by two-sided Wilcoxon rank-sum test in **a** and **b** and by two-sided Student's *t*-test in **c** ( $***P < 2.2 \times 10^{-16}$ ).

across all chromosomes except B05, with B04 showing the greatest average B-to-A exchange (9.72 Mb), and was consistent across all four groups (Fig. 3c and Supplementary Fig. 10), indicating asymmetric genome evolution in peanut.

To assess population structure in HE variation, we hierarchically clustered 586 accessions using HE values for homoeologous gene pairs. The resulting four clusters were consistent with the SNP-defined population groups for more than 90% of accessions (Supplementary Table 28). The four groups exhibited distinct HE patterns (Fig. 3d and Supplementary Fig. 11), indicating that HE variation was structured by population. Further analysis revealed that 723 HE genes were shared across all four groups, whereas 45, 31, 17 and 14 group-specific HE genes were identified in the PA, HH, Fa and Vu groups, respectively (Supplementary Table 29). Gene ontology enrichment analysis revealed that shared HEs were predominantly associated with developmental and reproductive processes, whereas population-specific HEs were enriched for cellular-level functions, including cytoskeletal activity, cell cycle regulation and energy metabolism (Extended Data Fig. 4a). Genomically, shared HEs were highly concentrated on chromosome B05 (41.1%), from which population-specific HEs were nearly absent, whereas population-specific HEs were markedly enriched on B06 and B07, showing strong chromosome- and population-dependent biases (Supplementary Fig. 12). Furthermore, population-specific HEs displayed pronounced chromosome- and population-dependent biases. HH-specific HEs were primarily concentrated on B06, PA-specific HEs on B07 and Fa-specific HEs on B04, whereas Vu-specific HEs were more broadly distributed (Extended Data Fig. 4b).

To investigate the evolutionary origins of these HEs, we compared cultivated peanut HEs with the closest wild tetraploid relative, *Arachis monticola*. Among 723 shared HEs in cultivated peanut, 540 (74.7%) were also present in *A. monticola* (Extended Data Fig. 4c and Supplementary Table 30), indicating that the majority of shared HEs probably originated early after polyploid formation and have become largely fixed. Notably, 22 of 107 population-specific HEs (20.6%) were detectable in *A. monticola* (Extended Data Fig. 4c and Supplementary Table 30), suggesting that a subset of these events represents ancestral variation predating domestication, which was subsequently redistributed across cultivated populations. The B-to-A exchange bias observed in cultivated peanuts was also present in *A. monticola* (Extended Data Fig. 4d,e and Supplementary Table 31), implying that subgenome asymmetry reflects an intrinsic property rather than solely artificial selection. Frequency analysis of population-specific HEs revealed that HH- and Vu-specific HEs were enriched in improved cultivars, consistent with selection during breeding, whereas Fa-specific HEs showed reduced frequencies, and PA-specific HEs remained largely stable (Extended Data Fig. 4f). Pathway annotation indicated that HEs enriched in improved cultivars were predominantly associated with stress response and organ development (Supplementary Table 32), suggesting a potential contribution to traits under selection. Compared with non-HE homoeologous gene pairs, HE-affected pairs showed higher protein sequence identity (99.37% versus 97.91%) and lower  $K_s$  (0.0069 versus 0.034) and  $K_a$  (0.0028 versus 0.0084) values, suggesting reduced subgenomic divergence associated with HE (Supplementary Fig. 13).

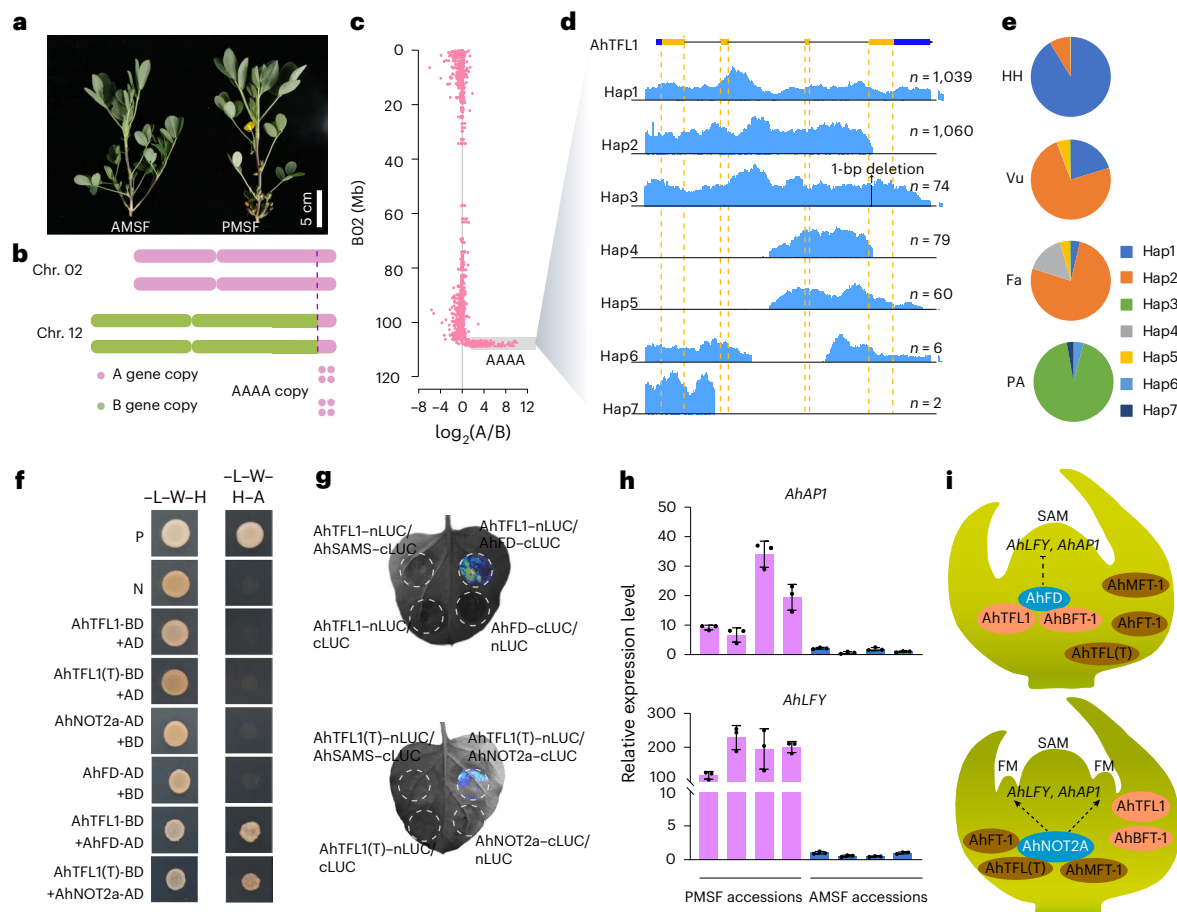


**Fig. 4 | Identification of the flowering regulator gene in peanut.** **a**, AMSF and PMSF in peanut. **b**, Duplication–deletion events caused by HE, illustrated with an AAAA genomic region. **c**, Distribution of HE events on chr. B02 based on  $\log_2(A/B)$  values. **d**, Seven haplotypes at the *AhTFL1* gene locus across 2,320 accessions. **e**, Distribution of seven *AhTFL1* haplotypes (Hap1 to Hap7) among the PA, HH, Vu and Fa groups. **f**, Yeast two-hybrid assay showing interactions of AhTFL1 and AhTFL1(T) proteins with other flowering regulators in peanut. Yeast was grown on SD/–Leu/–Trp/–His (–L–W–H) and SD/–Leu/–Trp/–His/–Ade (–L–W–H–A) media for assessment of protein–protein interactions. **g**, Split-luciferase

complementation imaging assay in *Nicotiana benthamiana* leaves. **h**, Relative expression levels of *AhAPI* and *AhLFY* in PMSF and AMSF peanut accessions. Four PMSF and four AMSF accessions were analyzed, with  $n = 3$  biological replicates per accession. Data are presented as the mean  $\pm$  s.d. **i**, Schematic model of the flowering pathway in peanut, illustrating the roles of peanut FT/TFL family in flower regulation in apical meristem (SAM) and floral meristem (FM) of shoots. In **f** and **g**, representative images from three independent experiments with similar results are shown.

## Pan-genome facilitates identification of agronomic genes hindered by HE

HE can induce duplication–deletion events (homoeologous replacements) and balanced translocations<sup>18,19</sup>, which can pose challenges for accurate gene identification when using a single genome as a ref.<sup>20</sup>. The genotypes of segmental allotetraploid regions are often missing owing to alignment of resequencing reads to both subgenomes; such reads are frequently discarded during variant calling (Supplementary Fig. 14). By contrast, the pan-genome approach, which incorporates broader species diversity, successfully mitigated these challenges. By using the pan-genome, we identified flowering regulation and dwarf genes in the HE regions with potential breeding value for improvement of plant architecture in peanut.

**Flowering regulator gene in peanut.** Presence or absence of flowers on the main stem (PMSF or AMSF, respectively) (Fig. 4a) is a major determinant of peanut architecture and pod setting<sup>3</sup>. Through pan-genome-based SV-GWAS, we identified a highly significant association signal within the HE region (AAAA copy region) at the distal end of chr. 12 (Fig. 4b,c and Extended Data Fig. 5a). The associated SV was a 1,489-bp deletion within the fourth exon of *AhTFL1* (*Ah12g463000*), which has previously been identified as the ortholog of *TFL1* in

*Arabidopsis*<sup>3,21,22</sup> (Extended Data Fig. 5b). Fine mapping independently localized the trait to a 117.6-kb interval containing 16 genes, including *AhTFL1* (Supplementary Fig. 15 and Supplementary Table 33). Together with previous evidence from *AhTFL1* mutant lines<sup>21</sup>, the concordant SV-GWAS and mapping results support *AhTFL1* as a candidate gene for main-stem flowering in peanut. Across 2,320 accessions, seven *AhTFL1* haplotypes were identified: the intact Hap1 was associated with AMSF, whereas structurally altered Hap2–Hap7 were associated with PMSF (Fig. 4d and Supplementary Note 2). Their frequencies varied among populations; Hap3 was enriched in PA, Hap1 in HH, and Hap2 in Vu and Fa, with Hap2 exhibiting the greatest haplotype diversity (Fig. 4e).

The Hap2-associated deletion generates a truncated AhTFL1 protein, AhTFL1(T) (Extended Data Fig. 5b). Phylogenetic analysis placed AhTFL1(T) within the FT rather than the TFL clade (Extended Data Fig. 5c). Transgenic experiments in *Arabidopsis* showed that 35S:*AhTFL1* expression delayed flowering, whereas 35S:*AhTFL1(T)* expression accelerated it (Extended Data Fig. 5d). Furthermore, 35S:*AhTFL1*-expressing plants had compact inflorescences filled with leaf primordia and indeterminate growth habits (Extended Data Fig. 5e,f). These results indicate that carboxy-terminal truncation switches AhTFL1 from a floral repressor to a flowering promoter. We further conducted a comprehensive analysis of the FT/TFL family in peanut (Extended Data Fig. 6a and

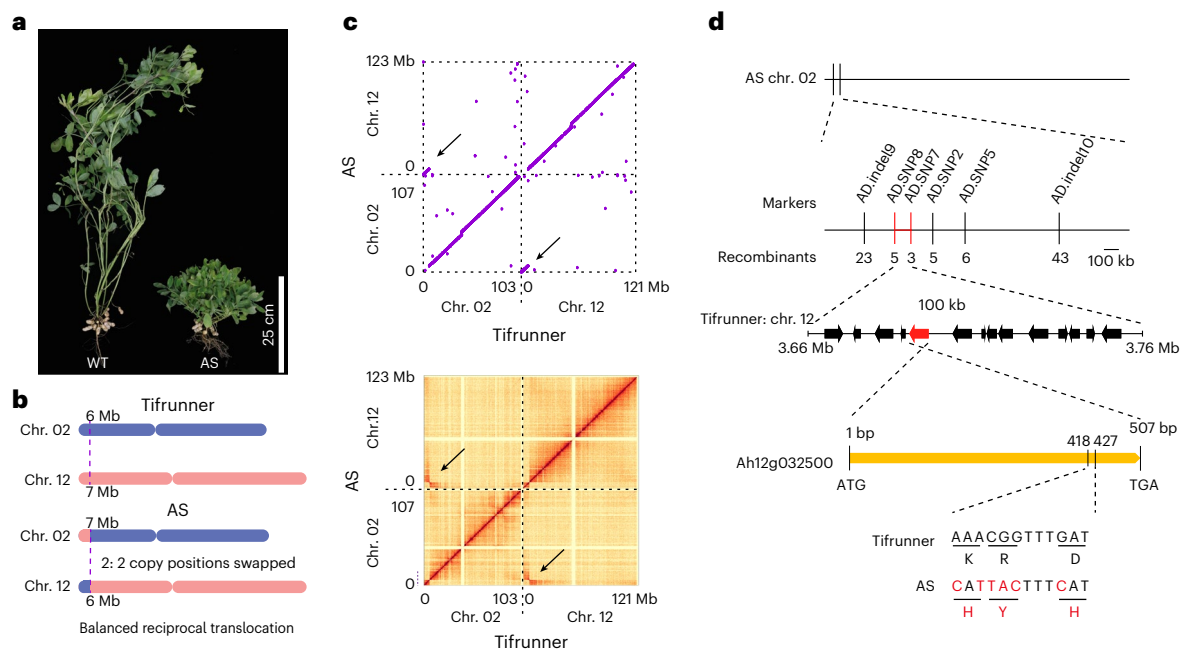


**Fig. 5 | Fine mapping of the dwarf gene in peanut. a**, Phenotype of AS compared to WT. **b**, Balanced reciprocal translocation caused by HE, illustrated with a 2:2 genomic configuration. **c**, Collinearity and Hi-C contact maps showing a

translocation of ~6 Mb identified at the end of the short arm of chr. 02 and chr. 12 in AS compared to the reference genome of Tifrunner. **d**, Fine mapping of the dwarf gene in peanut.

Supplementary Note 3). AhTFL1 and AhBFT-1, both of which delayed flowering in *Arabidopsis*, interacted with AhFD, whereas AhTFL1(T), AhFT-1 and AhMFT-1, all of which accelerated flowering, interacted with AhNOT2a (Fig. 4f,g and Extended Data Fig. 6b–d). Correspondingly, floral identity genes *AhAPI* and *AhLFY* were more highly expressed in PMSF than AMSF accessions (Fig. 4h). These findings suggest that AhTFL1 and its homologs play dual parts in control of flowering by modulating different protein interactions (Fig. 4i).

**Dwarf gene in peanut.** The green revolution represented by semidwarf and dwarf crop varieties has transformed modern agriculture<sup>21,23,24</sup>, but no dwarf germplasm has been well adopted in peanut cultivation. The accessions investigated in this study included the dwarf germplasm Aisheng (AS) belonging to var. *fastigiata*. AS plants present a highly compact phenotype, with a stem length 73.2% shorter than that of wild-type (WT) plants (Fig. 5a and Supplementary Table 34). Notably, no significant differences in key agronomic traits, such as hundred pod weight and hundred seed weight (Supplementary Table 34), were detected between AS and WT plants. Thus, AS represents a highly valuable genetic resource for development of high-yield dwarf cultivars in peanut.

To identify the gene underlying dwarfism in AS, we generated an  $F_2$  population of 1,548 individuals from a cross between AS and Tifrunner. The dwarf phenotype segregated as a single recessive nuclear locus (3:1;  $\chi^2 = 1.38$ ,  $P = 0.24$ ; Supplementary Table 35). Bulk-segregant analysis by deep sequencing and marker analysis initially mapped the locus to an interval of 3.0–5.0 Mb assigned to chromosome 2 in the Tifrunner reference. However, fine mapping revealed highly irregular recombination across this interval (Extended Data Fig. 7a–c). Genome alignment and Hi-C analysis identified an approximately 6.0-Mb balanced reciprocal translocation between chromosomes 2 and 12 in AS (Fig. 5b,c). The same translocation was detected in DL135 and Kaust by pan-genome analysis (Extended Data Fig. 7d and Supplementary Fig. 16), and a 3.94-kb HE region was identified at the translocation breakpoint (Supplementary Note 4). A marker derived from the breakpoint region identified the translocation in 463 of 2,320 accessions (Supplementary Fig. 17 and Supplementary Table 36). Of these, 432 accessions (93.30%) belonged to the Fa group, representing

89.44% of all Fa accessions (432 of 483). By contrast, the translocation was absent from PA (0 of 73) and rare in HH (1 of 935) and Vu (30 of 829). This distribution suggests that the translocation arose before or early during diversification of the Fa group and subsequently increased to high frequency. As DL135 carries the same translocation as AS, we crossed these accessions to restore structural collinearity during fine mapping. This population localized the dwarf locus to a 100-kb interval between AD.SNP7 and AD.SNP8 at 3.47–3.57 Mb on AS chromosome 2, corresponding to 3.66–3.76 Mb on Tifrunner chromosome 12 (Fig. 5d).

The interval contained 15 annotated genes and 23 sequence variants, comprising 6 indels and 17 SNPs. Six SNPs occurred within *Ah12g032500*, including three nonsynonymous substitutions (Fig. 5d and Supplementary Table 37). Screening of 2,320 accessions further identified an independent 47.31-kb deletion that disrupted *Ah12g032500* in DL156 (Extended Data Fig. 8a and Supplementary Fig. 18). DL156 exhibited dwarfism comparable to that of AS, with similar reductions in plant height and internode length (Extended Data Fig. 8b,c). Moreover, virus-induced silencing of *Ah12g032500* in WT peanut reduced both plant height and internode length relative to controls (Extended Data Fig. 8d–f), identifying *Ah12g032500* as a candidate gene underlying dwarfism. *Ah12g032500* encodes an LRR-containing protein that is highly conserved among wild *Arachis* relatives (>90% sequence identity) but substantially more diverged in other legumes (<50%; Supplementary Fig. 19). Gibberellin (GA) profiling revealed reduced levels of GA<sub>1</sub>, GA<sub>3</sub>, GA<sub>5</sub>, GA<sub>20</sub> and GA<sub>34</sub> in AS, whereas levels of the other measured phytohormones were unchanged (Extended Data Fig. 8g). Exogenous GA<sub>3</sub> partially restored shoot elongation and plant height in AS (Extended Data Fig. 8h,i). Together, these results suggest that *Ah12g032500* influences GA homeostasis in peanut.

### Pan-genome guided breeding for superior dwarf peanut lines

To identify superior haplotypes with potential breeding value, we analyzed phenotypic data for 25 traits, including key agronomic characteristics such as plant architecture, yield-related attributes and quality traits, across 555 accessions. Using both SNP- and SV-based GWAS within a graph-based pan-genome framework, we identified 221 loci (122 from the SV-GWAS and 99 from the SNP-GWAS) that were significantly

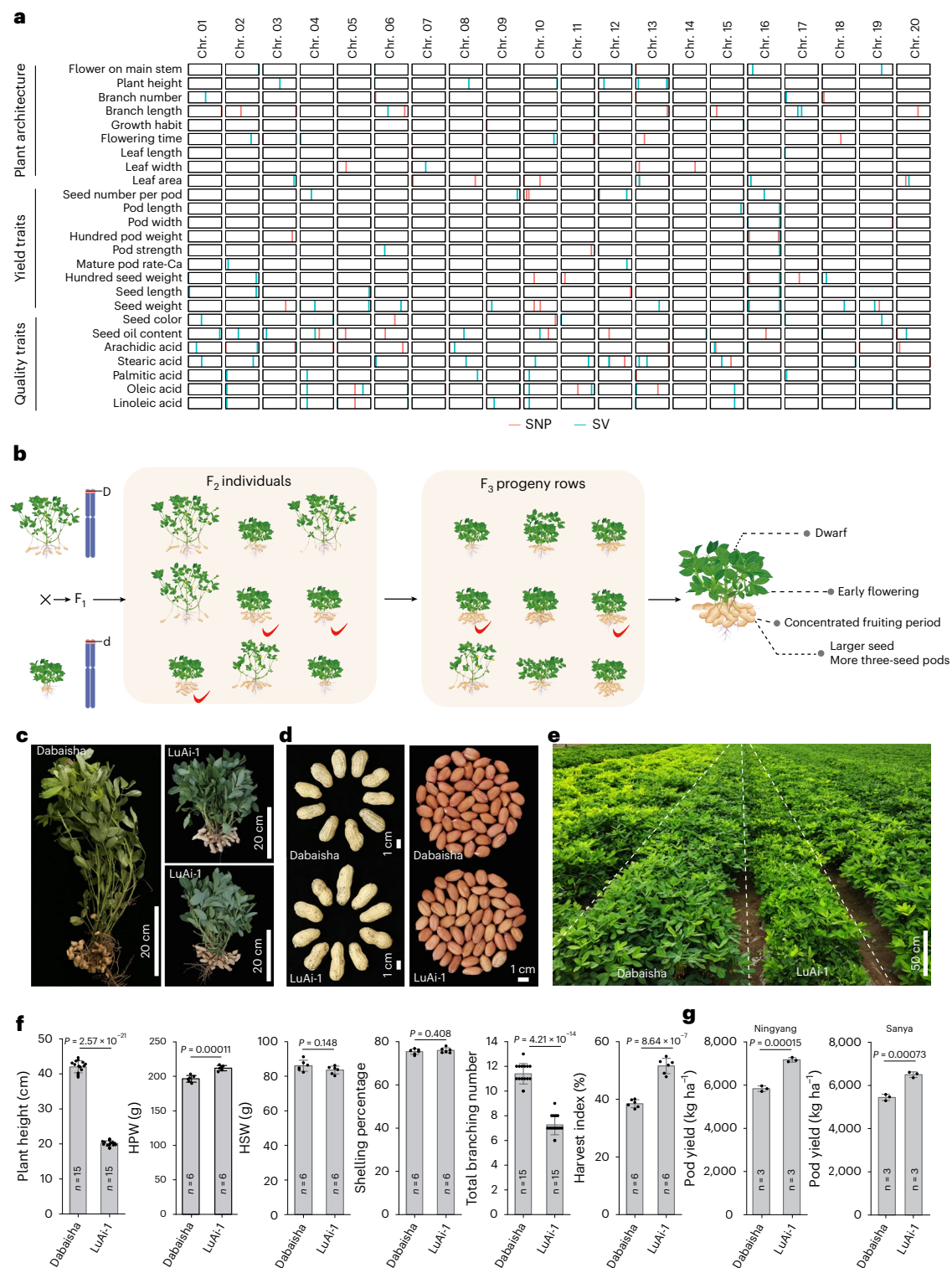


**Fig. 6 | Identification of loci related to important traits and application for high-yield cultivar selection. a**, Distribution of loci associated with agronomic traits across the whole genome. ‘Mature pod rate-Ca’ denotes the comparison between calcium and no-calcium conditions. **b**, Diagram showing the breeding process from crossing  $F_1$  individuals (left) to the selection of desired traits in  $F_3$  progeny rows (right). In the  $F_2$  generation, individuals displayed phenotypic variation with respect to traits such as dwarfism, early flowering and larger seeds. Selected progeny rows in the  $F_3$  generation exhibited a desirable combination of traits, including dwarfism, early flowering, concentrated fruiting period, larger seeds and an increased number of three-seed pods. **c**, Morphology of parental (Dabaisha) and improved (LuAi-1) lines. **d**, Pod and seed morphology of

Dabaisha (top row) and LuAi-1 (bottom row). **e**, Field performance of Dabaisha (left) and LuAi-1 (right). **f**, Comparison of agronomic traits between Dabaisha and LuAi-1. Plant height and total branching number were measured from  $n = 15$  biologically independent plants; hundred pod weight, hundred seed weight, shelling percentage and harvest index were measured from  $n = 6$  biologically independent samples. **g**, Field yield performance of Dabaisha and LuAi-1 at Ningyang in 2024 and Sanya in 2025 ( $n = 3$  independent field plots at each location). Data are presented as the mean  $\pm$  s.d. Statistical significance was assessed by two-sided Student’s  $t$ -test. HPW, hundred pod weight; HSW, hundred seed weight.

associated with these traits (Fig. 6a, Supplementary Figs. 20–27 and Supplementary Table 38). Linkage disequilibrium (LD) analysis showed that 107 loci (48.4%) were specific to SVs and 85 (38.5%) to SNPs, with only 29 (14 SNP loci and 15 SV loci; 13.1%) colocalizing within the same LD blocks (Supplementary Table 38). Notably, approximately 23% of these loci overlapped with those reported in previous studies, reinforcing the robustness of our GWAS findings. Further analysis of these loci based on haplotype–trait associations from GWAS identified 157 superior haplotypes associated with these traits, of which 83 (52.9%) were attributable to SVs and 74 (47.1%) to SNPs; these have potential value for breeding applications (Supplementary Table 39). Representative loci associated with seed size and plant architecture were further examined through functional assays (Supplementary Note 5 and Supplementary Figs. 28–30).

To develop high-yield, dwarf peanut lines with improved agronomic traits, we utilized AS as the genetic background, incorporating different subsets of superior haplotypes and population variation data. First, from among the identified materials with the same balanced translocations on chr. 02 and 12 as AS, we selected Dabaisha, a high-yield Chinese cultivar, as the breeding parent to cross with AS. We then performed genotyping analysis on the F<sub>2</sub> population from Dabaisha × AS, based on the superior haplotype dataset to identify individuals with desirable traits. The selected individuals were further propagated and advanced through generations (Supplementary Note 6). Using marker-assisted selection and phenotype-based selection, we successfully developed LuAi-1, a line that harbored favorable haplotypes for traits including dwarfism, early flowering, larger seed size, concentrated pod setting and a higher proportion of three-seed pods (Fig. 6b–d). Compared to the control Dabaisha, LuAi-1 exhibited a greater hundred pod weight, reduced plant height and fewer branches, with no significant differences in hundred seed weight (Fig. 6e,f). Notably, LuAi-1 had a higher harvest index than the control. In field trials, LuAi-1 demonstrated a yield increase ranging from 17.28% to 20.79% compared to Dabaisha under high-density planting conditions in 2 consecutive years (Supplementary Fig. 31). In addition, it exhibited a yield increase of 19.34% to 22.90% in large-scale field trials, further confirming its superior performance (Fig. 6g). In conclusion, our breeding strategy utilizing the dwarf gene to achieve plant height reduction enhances yield potential and sustainability in peanut production.

## Discussion

Advances in graph pan-genomes have led to the discovery of hidden structural variation in complex genomes, particularly in polyploids<sup>25–27</sup>. By integrating 14 reference-quality genome assemblies representing all 6 botanical varieties of cultivated peanut, we constructed a graph-based pan-genome and genotyped SVs across 2,320 globally representative accessions. The graph framework improved population-scale SV detection, particularly in HE-affected regions that were poorly represented by a single linear reference, thereby facilitating the identification of trait-associated variants and haplotypes.

In polyploid crops, HEs between subgenomic chromosomes represent a key mechanism for generating novel genetic variation<sup>28</sup>. In cultivated peanut, the strong concordance between HE-based clustering and SNP-defined population structure indicates that HE variation is nonrandom and may be shaped by population-level evolutionary forces. The consistent B-to-A exchange bias further suggests that postpolyploid genome stabilization followed an asymmetric trajectory. Similar directional biases have been reported in other allopolyploid crops, including wheat<sup>28</sup>, strawberry<sup>29</sup> and Brassica<sup>30,31</sup>, consistent with asymmetric subgenome evolution. In peanut, the B subgenome shows distinct transposable-element content and DNA methylation profiles relative to the A subgenome<sup>2,32</sup>, suggesting pre-existing regulatory asymmetry that may have contributed to preferential retention of B-derived segments. The shared B-to-A bias in *A. monticola* indicates that HE directionality predated domestication, whereas domestication and breeding subsequently reshaped the frequencies of specific exchanged segments.

Pan-genome analysis identified an HE-associated SV in *AhTFL1* linked to main-stem flowering, extending previous quantitative trait locus mapping and candidate gene studies in peanut<sup>21,22</sup>. *TFL1* homologs, including *Dt1* in soybean and *PvTFL1* in common bean, regulate stem determinacy and flowering time in legumes<sup>33–35</sup>. Our findings are consistent with a conserved role of *TFL1* homologs in legume flowering. In *Arabidopsis*, *TFL1* and *FT* act antagonistically through shared FD/FDP and 14-3-3 partners to regulate floral identity genes, including *LFY* and *API*<sup>36–39</sup>. In cucumber, *CsTFL1* competes with *CsFT* for interaction with *CsNOT2a*–*CsFDP* to inhibit determinate growth and terminal flower formation<sup>40</sup>. Peanut seems to use a related but distinct configuration: full-length *AhTFL1*, a floral repressor, interacted with *AhFD*, whereas flowering-promoting *AhTFL1(T)* interacted with *AhNOT2a*. Together, these findings support a model in which C-terminal truncation redirects *AhTFL1* partner specificity and converts it into an *FT*-like regulator that promotes main-stem flowering.

Mutations affecting GA metabolism or signaling underpinned the green revolution in crops<sup>41–43</sup>. In peanut, dwarf architecture facilitates mechanized, high-density production<sup>44,45</sup>. Using the peanut pan-genome, we identified *Ah12g032500* as a candidate gene for dwarfism. *Ah12g032500* resides in a 6.0-Mb balanced translocation between chromosomes 02 and 12, an SV that has hindered its genetic identification and breeding application. The presence of this translocation in Fa further indicates that HE-associated rearrangements contribute to genomic differentiation among peanut groups. *Ah12g032500* encodes an LRR-domain protein potentially involved in GA-related regulation. Similarly, GA-sensitive wheat dwarfing gene *Rht13* encodes an autoactive NB-LRR protein<sup>46,47</sup>, suggesting that LRR-domain proteins may represent a previously unrecognized class of regulators of GA-mediated plant height. These findings illustrate the utility of pan-genome analysis for gene discovery in structurally complex HE regions of allopolyploid genomes.

To translate these findings into breeding, we used AS as the genetic background and combined favorable haplotypes associated with main-stem flowering, increased seed size and three-seed pods through marker-assisted selection. This strategy generated LuAi-1, a dwarf high-yield line, thereby demonstrating the practical utility of pan-genome-informed breeding for development of peanut germplasm suited to dense planting and mechanized production.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-026-02765-x>.

## References

- Bertioli, 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).
- Bertioli, D. J. et al. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat. Genet.* **51**, 877–884 (2019).
- Krapovickas, A., Gregory, W. C., Williams, D. E. & Simpson, C. E. Taxonomy of the genus *Arachis* (Leguminosae). *Bonplandia* **16**, 7–205 (2007).
- Hammons, R. O., Herman, D. & Stalker, H. T. Origin and early history of the peanut. In *Peanuts: Genetics, Processing, and Utilization* (eds Stalker, H. T. & Wilson, R. F.) Ch. 1 (Academic Press and AOCS Press, 2016).
- Pandey, M. K. et al. Advances in *Arachis* genomics for peanut improvement. *Biotechnol. Adv.* **30**, 639–651 (2012).
- Barkley, N. A., Upadhyaya, H. D., Liao, B. & Holbrook, C. C. Global resources of genetic diversity in peanut. In *Peanuts: Genetics, Processing, and Utilization* (eds Stalker, H. T. & Wilson, R. F.) Ch. 3 (Academic Press and AOCS Press, 2016).

7. Raza, A. et al. Designing future peanut: the power of genomics-assisted breeding. *Theor. Appl. Genet.* **137**, 66 (2024).
8. Lu, Q. et al. A genomic variation map provides insights into peanut diversity in China and associations with 28 agronomic traits. *Nat. Genet.* **56**, 530–540 (2024).
9. Zhuang, W. et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* **51**, 865–876 (2019).
10. Driguez, P. et al. LeafGo: leaf to genome, a quick workflow to produce high-quality de novo plant genomes using long-read sequencing technology. *Genome Biol.* **22**, 256 (2021).
11. Newman, C. S. et al. Initiation of genomics-assisted breeding in Virginia-type peanuts through the generation of a *de novo* reference genome and informative markers. *Front. Plant Sci.* **13**, 1073542 (2023).
12. Zhao, K. et al. Pangenome analysis reveals structural variation associated with seed size and weight traits in peanut. *Nat. Genet.* **57**, 1250–1261 (2025).
13. Chen, Y. et al. Identification of major QTL for seed number per pod on chromosome A05 of tetraploid peanut (*Arachis hypogaea* L.). *Crop J.* **7**, 238–248 (2019).
14. Smartt, J. *The Groundnut Crop: A Scientific Basis for Improvement* (Springer Science & Business Media, 2012).
15. Lin, X. et al. Novel and multifaceted regulations of photoperiodic flowering by phytochrome A in soybean. *Proc. Natl Acad. Sci. USA* **119**, e2208708119 (2022).
16. Gualberti, G. et al. Mutations in the Dof zinc finger genes *DAG2* and *DAG1* influence with opposite effects the germination of *Arabidopsis* seeds. *Plant Cell* **14**, 1253–1263 (2002).
17. Leal-Bertioli, S. C. M. et al. Segmental allopolyploidy in action: increasing diversity through polyploid hybridization and homoeologous recombination. *Am. J. Bot.* **105**, 1053–1066 (2018).
18. Mason, A. S. & Wendel, J. F. Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. *Front. Genet.* **11**, 1014 (2020).
19. Deb, S. K., Edger, P. P., Pires, J. C. & McKain, M. R. Patterns, mechanisms, and consequences of homoeologous exchange in allopolyploid angiosperms: a genomic and epigenomic perspective. *New Phytol.* **238**, 2284–2304 (2023).
20. Clevenger, J. et al. Genome-wide SNP genotyping resolves signatures of selection and tetrasomic recombination in peanut. *Mol. Plant* **10**, 309–322 (2017).
21. Kunta, S. et al. Identification of a major locus for flowering pattern sheds light on plant architecture diversification in cultivated peanut. *Theor. Appl. Genet.* **135**, 1767–1777 (2022).
22. Zheng, Z. et al. Chloroplast and whole-genome sequencing shed light on the evolutionary history and phenotypic diversification of peanuts. *Nat. Genet.* **56**, 1975–1984 (2024).
23. Xue, H., Zhang, Y. & Xiao, G. Neo-gibberellin signaling: guiding the next generation of the Green Revolution. *Trends Plant Sci.* **25**, 520–522 (2020).
24. Liu, Q. et al. Improving crop nitrogen use efficiency toward sustainable Green Revolution. *Annu. Rev. Plant Biol.* **73**, 523–551 (2022).
25. Song, J. et al. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants* **6**, 34–45 (2020).
26. Jayakodi, M. et al. Structural variation in the pangenome of wild and domesticated barley. *Nature* **636**, 654–662 (2024).
27. Jiao, C. et al. Pan-genome bridges wheat structural variations with habitat and breeding. *Nature* **637**, 384–393 (2025).
28. Zhang, Z. et al. Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids. *Proc. Natl Acad. Sci. USA* **117**, 14561–14571 (2020).
29. Edger, P. P. et al. Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* **51**, 541–547 (2019).
30. Samans, B., Chalhouh, B. & Snowdon, R. J. Surviving a genome collision: genomic signatures of allopolyploidization in the recent crop species *Brassica napus*. *Plant Genome* **10**, 2012–2017 (2017).
31. Wang, T., van Dijk, A. D., Zhao, R., Bonnema, G. & Wang, X. Contribution of homoeologous exchange to domestication of polyploid *Brassica*. *Genome Biol.* **25**, 231 (2024).
32. Chen, X. et al. Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and oil improvement. *Mol. Plant* **12**, 920–934 (2019).
33. Liu, B. et al. The soybean stem growth habit gene *Dt1* is an ortholog of *Arabidopsis* *TERMINAL FLOWER1*. *Plant Physiol.* **153**, 198–210 (2010).
34. Tian, Z. et al. Artificial selection for determinate growth habit in soybean. *Proc. Natl Acad. Sci. USA* **107**, 8563–8568 (2010).
35. Repinski, S. L., Kwak, M. & Gepts, P. The common bean growth habit gene *PvTFL1y* is a functional homolog of *Arabidopsis* *TFL1*. *Theor. Appl. Genet.* **124**, 1539–1547 (2012).
36. Hanano, S. & Goto, K. *Arabidopsis* *TERMINAL FLOWER1* is involved in the regulation of flowering time and inflorescence development through transcriptional repression. *Plant Cell* **23**, 3172–3184 (2011).
37. Wickland, D. P. & Hanzawa, Y. The *FLOWERING LOCUS T/TERMINAL FLOWER 1* gene family: functional evolution and molecular mechanisms. *Mol. Plant* **8**, 983–997 (2015).
38. Abe, M. et al. FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* **309**, 1052–1056 (2005).
39. Taoka, K. et al. 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. *Nature* **476**, 332–335 (2011).
40. Wen, C. et al. CsTFL1 inhibits determinate growth and terminal flower formation through interaction with CsNOT2a in cucumber. *Development* **146**, dev180166 (2019).
41. Peng, J. et al. ‘Green revolution’ genes encode mutant gibberellin response modulators. *Nature* **400**, 256–261 (1999).
42. Sasaki, A. et al. A mutant gibberellin-synthesis gene in rice. *Nature* **416**, 701–702 (2002).
43. Spielmeier, W., Ellis, M. H. & Chandler, P. M. Semidwarf (*sd-1*), “green revolution” rice, contains a defective gibberellin 20-oxidase gene. *Proc. Natl Acad. Sci. USA* **99**, 9043–9048 (2002).
44. Giayetto, O., Cerioni, G. A. & Asnal, W. E. Effect of sowing spacing on vegetative growth, dry matter production, and peanut pod yield. *Peanut Sci.* **25**, 86–87 (1998).
45. Zhao, J. et al. Effects of paclobutrazol application on plant architecture, lodging resistance, photosynthetic characteristics, and peanut yield at different single-seed precise sowing densities. *Crop J.* **11**, 301–310 (2023).
46. Rebetzke, G. J. et al. The *Rht13* dwarfing gene reduces peduncle length and plant height to increase grain number and yield of wheat. *Field Crops Res.* **124**, 323–331 (2011).
47. Borrill, P. et al. An autoactive *NB-LRR* gene causes *Rht13* dwarfism in wheat. *Proc. Natl Acad. Sci. USA* **119**, e2209875119 (2022).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2026

<sup>1</sup>Institute of Crop Germplasm Resources, Shandong Academy of Agricultural Sciences/Shandong Provincial Key Laboratory of Field Crop Physiology, Ecology, and Efficient Production, Jinan, China. <sup>2</sup>Centre of Excellence in Genomics & Systems Biology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India. <sup>3</sup>WA State Agricultural Biotechnology Centre, Centre for Crop and Food Innovation, Food Futures Institute, Murdoch University, Murdoch, Western Australia, Australia. <sup>4</sup>Cash Crops Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, China. <sup>5</sup>RS Paroda Genebank, ICRISAT, Patancheru, India. <sup>6</sup>State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China. <sup>7</sup>Henan Academy of Agricultural Sciences, Zhengzhou, China. <sup>8</sup>These authors contributed equally: Yiyang Liu, Weitao Li, Rongchong Li, Manish K. Pandey. ✉e-mail: [ligw\\_saas@163.com](mailto:ligw_saas@163.com); [haasz@126.com](mailto:haasz@126.com); [rajeev.varshney@murdoch.edu.au](mailto:rajeev.varshney@murdoch.edu.au); [wanshubo2016@163.com](mailto:wanshubo2016@163.com)

## Methods

### Plant materials and sample collection

We selected 10 representative peanut accessions for de novo genome assembly, based on phylogenetic relationships of 401 accessions (PRJNA1163959; Supplementary Note 1 and Supplementary Fig. 1). Tissue samples including leaves, roots, flowers and stems were collected for transcriptomic analysis. For population analysis, we used a total of 2,320 accessions of cultivated peanut from 87 countries and regions, representing 6 botanical varieties of cultivated peanut. Genomic DNA was extracted from fresh seedling leaves collected from a single plant per accession. Samples for DNA or RNA extraction were immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . For the GWAS, 555 accessions were grown in 2019 and 2020 at the Ningyang ( $116.80^{\circ}\text{E}$ ,  $35.77^{\circ}\text{N}$ ) and Jinan ( $117.12^{\circ}\text{E}$ ,  $36.65^{\circ}\text{N}$ ) experimental stations of the Shandong Academy of Agricultural Sciences, China. Field trials used a randomized block design with two replicates, at least 30 plants per plot, 30-cm row spacing and 10-cm plant spacing. Crop management followed the China National Peanut Cultivar Field Test protocol. Twenty-five traits were measured in at least six plants per accession, and accession means were used for analysis.

### Genome assembly

For de novo genome assembly, we assembled HiFi reads to obtain a preliminary assembly based on the Highly Accurate Single-Molecule Consensus Reads using Hifiasm (v.0.14.2-r375)<sup>48</sup> with default parameters. Subsequently, Hi-C data were mapped to the assembled contigs using Juicer (v.1.6)<sup>49</sup>. The mapped valid reads were input to the 3D-DNA pipeline<sup>50</sup> for grouping, ordering and orientating of the contigs. For genome gap filling, the ONT reads were assembled using NextDenovo<sup>51</sup> (self-corrected parameters: `--reads_cutoff 5k --seed_depth 45 --seed_cutoff auto`; and default assembly parameters). The assembled contigs were aligned with the genome sequences using the unimap software (v.0.1-r46-dirty; <https://github.com/lh3/unimap>) to search for contigs that could cross the gap. The searched contigs were then used to fill the gap. Moreover, we used ONT reads to close the gap sequence of the genome by LR\_Gapcloser<sup>52</sup> with parameters `-s n -r 2`. Finally, we used the filtered clean second-generation data to polish the gap-filled genome using NextPolish for three rounds<sup>53</sup>.

### Gene prediction and TE annotation

To annotate TEs and predict genes, we first constructed a de novo repeat library using the Extensive de novo TE Annotator (v.2.1.3)<sup>54</sup>, employing parameters `--sensitive 1 --anno 1`. TEs were subsequently identified using RepeatMasker (v.4.07, <http://repeatmasker.org/>) with the option `-no_is`.

For protein-coding gene annotation, the protein sequences of *Glycine max*, *Arabidopsis thaliana*, *Vitis vinifera*, *A. duranensis*, *A. ipaensis*, *Arachis hypogaea* cultivar Tifrunner and *A. monticola* from Phytozome (<https://phytozome.jgi.doe.gov>) and PeanutBase (<https://www.peanutbase.org/data/public>) were used as homologous protein evidence. Transcriptome evidence was generated in this study from RNA-seq datasets for each accession, encompassing root, stem, leaf, flower and seed tissues. The RNA-seq reads were mapped to the corresponding genome using HISAT2 (v.2.1.0)<sup>55</sup>. Full-length transcript sequences were then assembled with StringTie2 (v.2.1.7)<sup>56</sup> using default parameters, and these assembled transcripts were subsequently utilized as transcriptomic evidence. Program to Assemble Spliced Alignments (PASA) v.2.4.1<sup>57</sup> was used to annotate the genomic structure based on transcript evidence with default parameters.

The MAKER2 pipeline (v.2.31.11)<sup>58</sup> was used to integrate ab initio predictions, homology-based predictions and transcriptome evidence. In brief, the following steps were performed: (1) RepeatMasker (v.4.07) was used to mask repetitive regions in the genome assembly before gene prediction; (2) ab initio gene prediction was carried out using AUGUSTUS (v.3.4.0)<sup>59</sup>; (3) BLASTN was used to align the transcript

evidence to the repeat-masked genome, BLASTX to align the homologous protein evidence to the genome and Exonerate (v.2.2.0)<sup>60</sup> to realign the BLAST hits to the genome; and (4) the predicted gene models were integrated using MAKER2 based on hints obtained from the above alignments.

We then used EvidenceModeler v.1.1.1<sup>57</sup> to integrate the results from MAKER2 and PASA and generate nonredundant annotations. To ensure exclusion of TE coding regions, TE protein domains were identified using TESorter<sup>61</sup> and removed during the EvidenceModeler integration. Finally, PASA was used to generate information about untranslated regions and alternative splicing in final gene models.

For functional annotation, the predicted gene models were annotated using EggNOG-mapper (v.1.0.3)<sup>62</sup> based on the EggNOG database. In addition, protein-coding regions were aligned to public databases, including NR (<https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/>) and TrEMBL (<https://www.ebi.ac.uk/uniprot>), using DIAMOND (v.0.9.24.125)<sup>63</sup> with an *E*-value threshold of  $\leq 1.0 \times 10^{-5}$ .

### Genome quality evaluation

The quality of the 10 genome assemblies was assessed by the following methods. Short (pair-end) reads were remapped to the final assembled genome to assess completeness using BWA-MEM (v.0.7.17)<sup>64</sup>. HiFi and ONT long reads were aligned to the genome assemblies using minimap2 (v.2.21-r1071)<sup>65</sup> to validate the accuracy of genome assemblies. Alignment metrics, including mapping rates and coverage, were calculated to assess the assembly quality and accuracy. BUSCO<sup>66</sup> v.5.5.0 with the embryophyta\_odb10 gene set was further used to assess the completeness of the assembled genome. The LTR assembly index was used to evaluate assembly continuity based on full-length LTR retrotransposons using LTR\_retriever (v.3.0.1)<sup>67</sup>. *K*-mer-based completeness, quality and false duplication evaluation were performed with Merqury (v.1.3)<sup>68</sup>.

### Gene-based pan-genomic analysis

The ten newly assembled genomes and four previously published genomes (Tifrunner<sup>2</sup>, Shitouqi<sup>9</sup>, Kaust<sup>10</sup> and Bailey<sup>11</sup>) were used to construct the gene-based pan-genome. To normalize the gene predictions across the 14 *A. hypogaea* genomes, we adjusted their annotation results using a cross-mapping method, following strategies reported previously<sup>26,27</sup>. First, we performed pairwise collinearity analysis for the 14 assemblies using MCScanX<sup>69</sup> with default parameters to guide construction of a unified gene set. We then used the Tifrunner genes as the initial reference set and incrementally incorporated genes from the other 13 assemblies. A gene was added to the set only if it did not have a collinear counterpart in the gene set generated in the previous step. This iterative process continued until all genes from the 13 assemblies had been included. Subsequently, we remapped all genes from this unified gene set to the 14 assemblies using GMAP<sup>70</sup> with the following specific parameters: `--format=gff3 --gene -n 10 --min-trimmed-coverage=90 --min-identity=90`. Gene models identified through remapping that did not overlap with any existing gene models were incorporated into the individual genome annotations. To avoid inflation of private gene numbers due to misannotations in individual de novo genome assemblies, we filtered the gene set using BLASTP against protein datasets from *A. duranensis*, *A. ipaensis*, *Vigna radiata*, *G. max*, *Lotus corniculatus*, *Medicago truncatula*, *Oryza sativa*, and *A. thaliana* (downloaded from Phytozome, <https://phytozome.jgi.doe.gov>), with an *e*-value cutoff of  $1 \times 10^{-10}$ . Genes were retained only if they showed significant similarity to at least one reference protein, with similarity thresholds of >90% for ancestral *Arachis*, >70% for Fabaceae relatives and >60% for distant angiosperm models (*O. sativa* and *A. thaliana*).

The adjusted protein sequences in each peanut genome were clustered into gene families using orthofinder (v.2.5.5)<sup>71</sup> with the following parameters: `-t 60 -a 60 -M msa -I iqtree -ot`. The resulting gene families were classified into core, softcore, dispensable and private categories according to their presence in each peanut genome. A gene family was

defined as a core gene family if it was present in all 14 genomes, as a softcore gene family if it was present in 13 genomes, as a dispensable gene family if it was present in 2–12 genomes or as a private gene family if it was present in 1 genome. The ratio of nonsynonymous/synonymous substitution ( $K_a/K_s$ ) was calculated using ParaAT.pl (v.2.0)<sup>72</sup> with the following parameters: -p proc -g -k -f axt. The ParaAT.pl program was based on KaKs\_Calculator (v.2.0)<sup>73</sup>.

### Identification of genetic variants

Genetic variants were identified using minimap2<sup>65</sup> and SyRI (v.1.5.4)<sup>74</sup>. SNPs, small indels ( $\leq 50$  bp), PAVs ( $> 50$  bp), inversions and translocations were detected by alignment of the chromosomes of the 13 genomes to the *A. hypogaea* Tifrunner v.2 reference genome. Chromosome alignments were performed using minimap2 with parameters -ax asm5 -eqx, followed by SyRI analysis with parameters -k -F B for detection of SVs.

For CNVs, the chromosomes of the 13 genomes were first aligned to the Tifrunner v.2 genome using nucmer (MUMmer v.4.0.0rc1)<sup>75</sup> with default parameters. CNVs were then identified based on the resulting alignments using SVMU<sup>76</sup>. SNPs and small indels identified across the 13 genomes were merged using vcf-merge from VCFtools (v.0.1.17)<sup>77</sup>. Given the identified PAVs, inversions, translocations and CNVs, the SURVIVOR (v.1.0.7)<sup>78</sup> tool was used with merging parameters 50 1 0 0 0 0 to obtain nonredundant SVs. Syntenic alignments, inversions, translocations and duplications were visualized using plotsr (v.1.1.1)<sup>79</sup> with the following parameters: --sr -W 8 -H 4.

### Graph-based pan-genome construction

To better explore the peanut genomic variations, we constructed a graph-based pan-genome using vg toolkit (v.1.49.0)<sup>80</sup>. The reference sequences of Tifrunner, merged SNPs, insertions, deletions, inversions, translocations and CNVs were constructed into a pan-genome using vg with the following parameters: construct -a -f -S. The xg index and dist index were generated using vg index with parameters -x and -j, respectively. The gbwt index was generated using vg gbwt with parameters -g -x -P, and the min index was built using vg minimizer with parameters -g -d.

### Variant calling of 2,320 accessions

For SNP and small indel calling, low-quality sequencing reads from the 2,320 accessions were removed using fastp (v.0.23.0)<sup>81</sup> with default parameters. The filtered reads were then aligned to the constructed pan-genome using vg giraffe<sup>82</sup> with parameters -o bam -p -b fast -m -d --gbwt-name -x, to produce alignments in BAM format. The BAM files were subsequently sorted and indexed using SAMtools sort and SAMtools index<sup>83</sup> with default settings. Small-variant discovery was performed using the Genome Analysis Toolkit (GATK) HaplotypeCaller (v.4.4.0.0)<sup>84</sup>; this generated individual GVCF files for each of the 2,320 accessions. These GVCF files were then combined into a single VCF file using GATK CombineGVCFs with default parameters. Genotyping was carried out using GATK GenotypeGVCFs to produce the final small-variant set. The VCF files from all 2,320 accessions were merged into a single VCF file for downstream analysis.

To identify structural variation in the 2,320 accessions, we mapped filtered short reads to the graph-based reference genome and genotyped SVs using vg giraffe<sup>82</sup> with parameters -o gaf -p -b fast -m -d --gbwt-name -x to generate alignments in GAF format. The pack file was generated using vg pack with parameters -x -a -Q 5, and the final SV results in vcf format were generated using vg call with parameters -r -k -a -A.

### Population analysis

We first filtered the raw SNPs using GATK Variant Filtration with parameters  $QD < 2.0 \parallel MQ < 40.0 \parallel FS > 60.0 \parallel SOR > 3.0 \parallel MQRankSum < -12.5 \parallel ReadPosRankSum < -8.0$  to remove low-quality variants.

The remaining variants were further filtered using PLINK (v.1.90p)<sup>85</sup> with parameters --indep-pairwise 1000 100 0.2. A 1 – identity-by-state genetic distance matrix was then calculated with PLINK using parameter --distance 1-ibs. Based on this genetic distance matrix, we constructed a neighbor-joining phylogenetic tree of 2,320 accessions using PHYLIP (v.3.69, <https://phylipweb.github.io/phylip/>). The resulting Newick tree was visualized using FigTree (v.1.4.3, <https://github.com/rambaut/figtree>). In addition, principal component analysis was performed using PLINK with parameter --pca.

To analyze population structure, we filtered the raw SNPs using VCFtools (v.0.1.17)<sup>77</sup> with parameters --maf 0.05 --max-missing 0.8. The filtered variants were then annotated using SnpEff (v.5.0)<sup>86</sup>, and 12,816 4Dtv sites were extracted from the annotated results. A BED format file was generated using PLINK with parameters --indep-pairwise 500 50 0.1 --make-bed --allow-extra-chr. The population structure information was evaluated using ADMIXTURE (v.1.3.0)<sup>87</sup> based on 12,816 4Dtv sites and plotted with Evolveview v.2<sup>88</sup> (<https://www.evogenius.info/evolveview-v2/#login>). To identify the best genetic cluster,  $K$ , we tested  $K$  values from 2 to 10 with 1,000 iterations for each run.  $K = 5$  was chosen as the optimal value. To assess genetic diversity and population differentiation, we calculated  $F_{ST}$  and nucleotide diversity ( $\pi$ ) values using VCFtools (v.0.1.17)<sup>77</sup> based on SNPs that had been filtered using GATK hard-filtering (v.4.4.0.0)<sup>84</sup>. The window size and step size were not set for the evaluation of  $F_{ST}$ , whereas  $\pi$  was calculated using parameters --window-pi 10000 --window-pi-step 5000.

### Case-control GWAS analysis

We implemented a case-control GWAS to identify SVs associated with distinct morphotypes of the two subspecies of cultivated peanut (*A. hypogaea*)<sup>89</sup>. Specifically, accessions from ssp. *hypogaea* were assigned a score of 1 (case group), and those from ssp. *fastigiata* were assigned a score of 0 (control group). This binary scoring approach was used to capture the phenotypic divergence between the two subspecies. We analyzed PAVs, which represent a major class of structural variation, to identify associations between genotypic variations and phenotypic divergence. Statistical significance was tested using two-sided Fisher's exact test and adjusted with Bonferroni correction. PAVs demonstrating significant associations were further annotated using BEDTools<sup>90</sup>; this enabled precise identification of candidate genes potentially driving the observed phenotypic differences between subspecies.

### Detection of HE

To assess HE genes, we analyzed the mapping depth of resequenced reads across 17,373 high-confidence homoeologous gene pairs from *A. duranensis* V14167 (A genome) and *A. ipaensis* K30076 (B genome), according to previous publications<sup>1,2</sup>. Briefly, after quality filtering, the resequencing data for each accession were aligned to the combined chromosomal pseudomolecule sequences of A genome and B genome using BWA-MEM (v.0.7.17)<sup>64</sup> with default parameters. The mean mapping depth of each gene was calculated using SAMtools coverage (v.1.11)<sup>83</sup> with default parameters. We determined the HE value for each gene pair by calculating the  $\log_2$  ratio of the mapping depth of the gene in the A genome to that of the corresponding gene in the B genome. A homoeologous gene pair was classified as an HE gene if the absolute value of the HE ratio exceeded 1.5. Specifically, a positive HE value indicated a preference for the A genome (AAAA), whereas a negative value indicated a preference for the B genome (BBBB). To mitigate potential bias introduced by local gene duplication or loss, we cross-verified HE values deviating from zero against the mapping depth of both homoeologous genes following the approach of Wang et al.<sup>31</sup>. An HE value was maintained if the mean mapping depth of 1 homoeologous gene exceeded 1.5 times the overall read depth in the accession and simultaneously that of the other gene was less than 0.525 times the overall read depth. Otherwise, the HE value was reset to zero (Supplementary Fig. 32).

To assess the HE regions in each accession, we merged adjacent or closely located HE genes with intergenic distances of less than 1 Mb into a single HE region. For each merged region, we recorded the starting position of the first HE gene and the ending position of the last HE gene to establish the boundaries. Isolated HE genes that exceeded the distance threshold were treated as separate HE regions. The length of each HE region was calculated as the difference between its start and end positions, corresponding to the genomic locations of the respective genes. Finally, we summarized the total number of HE regions, their cumulative lengths and their distribution across the genome for further analysis.

To evaluate HE at the group level, we first calculated a group-level mean HE value for each homoeologous gene pair, defined as the average  $\log_2(A/B)$  across all accessions within each group (PA, HH, Vu and Fa). Homoeologous gene pairs with group-level absolute mean HE > 1.5 were classified as group-level HE genes. Homoeologous gene pairs exceeding this threshold in all four groups were defined as shared HE genes, whereas those exceeding the threshold in only one group were designated group-specific HE genes.

To identify HE in *A. monticola*, we obtained eight publicly available resequencing datasets from a previous study<sup>12</sup>. Using the approach described above, we calculated HE values for each homoeologous gene pair in each individual. The HE values were then averaged across the eight samples to obtain a final group-level HE value for each gene pair. Subsequent analyses of HE events and HE regions were performed following the procedures applied to cultivated peanut accessions.

### Generation of transgenic plants

The coding sequences of *AhTFL1*, *AhTFL1(T)*, *AhBFT-1*, *AhMFT* and *AhFT-1* were cloned into the pHB vector<sup>91</sup>. Transgenic plants were then created using the floral-dip method<sup>92</sup>. The homozygous lines of transgenic plants at T3 generation were used for analysis in this study. Information about the primers used in this study is included in Supplementary Table 40.

### Yeast two-hybrid assays

Yeast two-hybrid assays were conducted using the Matchmaker Gold Two-Hybrid System (Clontech) following the manufacturer's protocol. The coding sequences of *AhTFL1*, *AhTFL1(T)*, *AhBFT-1*, *AhMFT* and *AhFT-1* were cloned into the pGBKT7 vector, whereas those of *AhFD* and *AhNOT2a* were cloned into the pGADT7 vector. Protein–protein interactions were assessed by cotransformation of the bait and prey plasmids into yeast strain Y2HGold. Yeast colonies were then grown on selective media (SD/–Leu/–Trp/–His and SD/–Leu/–Trp/–His/–Ade) and incubated at 30 °C for 5 days for evaluation of interaction strength.

### Split-luciferase complementation assay

The coding sequences of *AhTFL1*, *AhTFL1(T)*, *AhBFT-1*, *AhMFT* and *AhFT-1* were cloned into pCambia1300-35S-NLuc, whereas those of *AhFD* and *AhNOT2a* were cloned into pCambia1300-35S-CLuc. The constructs were introduced into *Agrobacterium tumefaciens* GV3101 and infiltrated into young, fully expanded leaves at a final optical density at 600 nm of 1.0 in infiltration buffer containing 10 mM MES (pH 5.6), 10 mM MgCl<sub>2</sub> and 150  $\mu$ M acetosyringone<sup>93</sup>. After incubation for 3 days at 28 °C under a 16-h light/8-h dark photoperiod, luciferase signals were imaged using a NightOWL III LB 983 low-light cooled CCD imaging system.

### Positional cloning

To identify the gene responsible for the dwarf trait in AS, we developed an F<sub>2</sub> segregating population from crosses between AS and DL135, which have the same balanced translocation in the target gene region. For fine mapping, indel and SNP markers were developed in the 3.0–5.0 Mb region on chr. A02 based on resequencing data from the parental lines. Using 6 markers within this interval and 267 recessive F<sub>2</sub> plants

exhibiting the typical dwarf phenotype, we successfully narrowed the locus down to a region of ~100 kb between two SNP markers, AD.SNP7 and AD.SNP8. Comparison of the 100-kb region between AS and DL135 revealed 23 variant sites; only 6 adjacent variants were located within the gene *Ah12g032500*, which encodes a leucine-rich repeat domain protein, with the others located in intergenic regions. These variants resulted in amino acid changes, confirming *Ah12g032500* as the target gene responsible for the dwarf phenotype. Detailed information for all markers used in the fine mapping is provided in Supplementary Table 40.

### SNP-based and SV-based GWAS

GWAS for 25 traits were performed separately using 1,356,521 SNPs and 18,542 SVs in the rMVP software package (1.0.8)<sup>94</sup>, employing a mixed linear model<sup>95</sup> to account for population structure and kinship. Variants were filtered for minor allele frequency  $\geq 0.05$  and missingness  $\leq 0.8$  for SNPs or missingness  $\leq 0.5$  for SVs. Genome-wide significance was determined using Bonferroni correction ( $P < 0.05/n$ , where  $n$  is the number of tested variants). Significant variants from SNP-GWAS were functionally annotated using SnpEff (v.5.0)<sup>86</sup>, whereas significant SVs were annotated using BEDTools<sup>90</sup> intersect, with parameters ‘-a -b -wa -wb’ applied for genomic overlap analysis. LD blocks were calculated using PLINK (v.1.90p)<sup>85</sup> with parameters ‘--blocks no-pheno-req --blocks-max-kb 600’. SNP and SV lead variants were considered to represent the same locus when they occurred within the same LD block.

### Identification of superior haplotypes

Superior genotypes were defined based on haplotype-level phenotypic performance for specific traits as previously described<sup>96</sup>. For haplotype analysis, we filtered the significant SNPs and/or SVs associated with 25 agronomic traits based on the following criteria: (i) minor allele frequency > 0.1; and (ii) proportion of missing calls per SNP and/or SV < 30%. The haplotypes within trait-associated loci were examined, and only homozygous genotypes were included in the subsequent analysis.

For the haplo-phenotype analysis, accessions were categorized based on their haplotype groups. Superior haplotypes were defined as those associated with the most favorable phenotypic traits (for example, higher 100-pod weight, 100-seed weight and seed number per pod). Specifically, haplotype-wise mean values for each trait were calculated by integrating the phenotypic data, and Duncan's multiple range test was used to assess statistical significance.

### Development and yield evaluation of LuAi-1

LuAi-1 was developed at Jinan, China (117.12° E, 36.65° N), from a cross between Dabaisha and AS. Approximately 200 dwarf F<sub>2</sub> plants were phenotyped and genotyped, and 40 individuals with favorable agronomic traits were selected and selfed through the F<sub>3</sub> generation. Two phenotypically uniform lines were selected to establish LuAi-1. Yield trials were conducted from May to September in 2023 and 2024 at Ningyang and Jinan using a randomized complete block design with three replicates and 15-m<sup>2</sup> plots. Large-scale trials were conducted at Ningyang from May to September 2024 and at Sanya (109.30° E, 18.15° N) from October 2024 to February 2025, with three replicates and 666.7-m<sup>2</sup> plots. Dabaisha was used as the control. Seeds were precision-sown at a target planting density of 300,000 plants per hectare, and standard agronomic practices were applied uniformly across treatments. At maturity, hundred pod weight, hundred seed weight, seed length and width, shelling percentage, plant height, branch number and harvest index were measured. Dry pod yield was recorded per plot and converted to yield per hectare.

### Statistics

Statistical analyses were performed in R (v.4.2.1). Two-sided Student's *t*-test was used for comparisons between two groups when data met the

assumptions of normality; otherwise, two-sided Wilcoxon rank-sum test was applied. One-way analysis of variance, as appropriate, was followed by Tukey's multiple-comparison test using the agricolae package in R. Case-control GWAS were performed using two-sided Fisher's exact test. Two-sided Pearson's chi-squared tests were used to compare frequencies of categorical variables among population groups. Exact statistical tests, sample sizes and *P* values are provided in the corresponding figure legends or source data, where applicable.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The genome assemblies of the 10 cultivated peanut accessions have been deposited at the National Genomics Data Center (NGDC) under accession [PRJCA070711](https://ngdc.cncr.ac.cn/data/PRJCA070711). The corresponding HiFi, ONT and Hi-C sequencing data have been deposited at NCBI under BioProjects [PRJNA1170232](https://ncbi.nlm.nih.gov/bioproject/PRJNA1170232), [PRJNA1169541](https://ncbi.nlm.nih.gov/bioproject/PRJNA1169541) and [PRJNA1170953](https://ncbi.nlm.nih.gov/bioproject/PRJNA1170953), respectively. Transcriptome data from different tissues of the 10 accessions of cultivated peanut can be found under BioProject [PRJNA1167437](https://ncbi.nlm.nih.gov/bioproject/PRJNA1167437). The results of the whole-genome sequencing of 2,320 accessions have been deposited at NCBI under BioProjects [PRJNA1163959](https://ncbi.nlm.nih.gov/bioproject/PRJNA1163959) and [PRJNA1071845](https://ncbi.nlm.nih.gov/bioproject/PRJNA1071845). Source data are provided with this paper.

### Code availability

The code used in this project is available via GitHub at <https://github.com/lrcchina/Peanut-pangenome> and Zenodo at <https://doi.org/10.5281/zenodo.13906381> (ref. 97).

### References

48. Cheng, H. Y., Concepcion, G. T., Feng, X. W., Zhang, H. W. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170 (2021).
49. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
50. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
51. Hu, J. et al. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol.* **25**, 107 (2024).
52. Xu, G. et al. LR\_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience*. **8**, giy157 (2019).
53. Hu, J., Fan, J., Sun, Z., Liu, S. & Berger, B. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255 (2020).
54. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).
55. Kim, D., Landmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 121–135 (2015).
56. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290 (2015).
57. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
58. Cantarel, B. L. et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
59. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
60. Slater, G. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
61. Zhang, R. G. et al. TESorter: an accurate and fast method to classify LTR-retrotransposons in plant genomes. *Hortic. Res.* **9**, uhac017 (2022).
62. Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
63. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
64. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
65. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
66. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
67. Ou, S., Chen, J. & Jiang, N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* **46**, e126 (2018).
68. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
69. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
70. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
71. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
72. Zhang, Z. et al. ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* **419**, 779–781 (2012).
73. Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. KaKs\_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinform.* **8**, 77–80 (2010).
74. Goel, M., Sun, H., Jiao, W. & Schneeberger, K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 277 (2019).
75. Marçais, G. et al. MUMmer4: a fast and versatile genome alignment system. *PLOS Comput. Biol.* **14**, e1005944 (2018).
76. Chakraborty, M., Emerson, J. J., Macdonald, S. J. & Long, A. D. Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. *Nat. Commun.* **10**, 4872 (2019).
77. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
78. Jeffares, D. C. et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat. Commun.* **8**, 14061 (2017).
79. Goel, M., Schneeberger, K. & Robinson, P. plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics* **38**, 2922–2926 (2022).
80. Garrison, E. et al. Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat. Biotechnol.* **36**, 875 (2018).
81. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
82. Sirén, J. et al. Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science* **374**, abg8871 (2021).

83. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
84. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
85. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
86. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2014).
87. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
88. He, Z. et al. Evolvview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res.* **44**, W236–W241 (2016).
89. Li, X. et al. Large-scale gene expression alterations introduced by structural variation drive morphotype diversification in *Brassica oleracea*. *Nat. Genet.* **56**, 517–529 (2024).
90. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
91. Liu, Y. et al. Genome-wide identification and expression of SAUR gene family in peanut (*Arachis hypogaea* L.) and functional identification of AhSAUR3 in drought tolerance. *BMC Plant Biol.* **22**, 178 (2022).
92. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
93. Wang, L., Yu, G., Macho, A. P. & Lozano-Durán, R. Split-luciferase complementation imaging assay to study protein-protein interactions in *Nicotiana benthamiana*. *Bio Protoc.* **11**, e4237 (2021).
94. Yin, L. et al. rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics Proteomics Bioinform.* **19**, 619–628 (2021).
95. Yu, J. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**, 203–208 (2005).
96. Varshney, R. K. et al. A chickpea genetic variation map based on the sequencing of 3,366 genomes. *Nature* **599**, 622–627 (2021).
97. Liu, Y. Pan-genome based resequencing of 2,320 accessions reveals structural variations and drives breeding advances in cultivated peanut. *Zenodo* <https://doi.org/10.5281/zenodo.13906381> (2024).

## Acknowledgements

We thank the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) and the United States Department of Agriculture (USDA) for providing the peanut materials. We also thank Z. Ren

(Shandong Agricultural University) for guidance in the positional cloning of genes in this study. We are grateful to former colleagues at the ICRISAT Genebank, particularly H. D. Upadhyaya and S. Singh, for their contributions during the initial phase of the study, including germplasm selection and field sowing for the collection of leaf samples for DNA isolation.

## Author contributions

G.L., Y. Liu, S.W., X.Z. and R.K.V. conceived and designed the research. R.K.V., M.K.P., A.C., Y. Liu, W.L. and G.L. prepared the materials and performed data analysis. A.C., R.L., V.G. and M.K.P. coordinated the genome sequencing and SV analysis. R.L., W.L. and R.R.M. performed the GWAS analysis. Y. Liu, W.L., G.T., L.L., L.Z. and Y. Li performed gene cloning and the molecular work on transgenic plants. B.B., Y.H., L.Z., S.L., F.D., F.C., L.S., F.X., P.X., L.H., R.T., F.G., X.L., J.Z., Q.H., Z.Z. and K.S. contributed to various analyses and interpretation. G.L., Y. Liu, W.L. and R.L. wrote and revised the paper, and G.L., X.Z., R.K.V. and S.W. finalized it. All authors read and approved the paper.

## Funding

This work was supported by the National Key Research and Development Program (2023YFD1202800), the National Natural Science Foundation of China (32272195 and 32472116), the Key Research and Development Program of Shandong province (2022LZGC021, 2021LZGC026, 2021LZGC025, 2022LZGCQY014), Taishan Scholar Program of Shandong province (tsqn202312285, tspd20221107), the project of ‘New Colleges and University 20’ of Jinan City (202333033), the Peanut Genomics of Afaltoxin project funded by MARS Inc. at ICRISAT and Murdoch University, and a Start Up grant to R.K.V. by the Food Futures Institute at Murdoch University.

## Competing interests

The authors declare no competing interests.

## Additional information

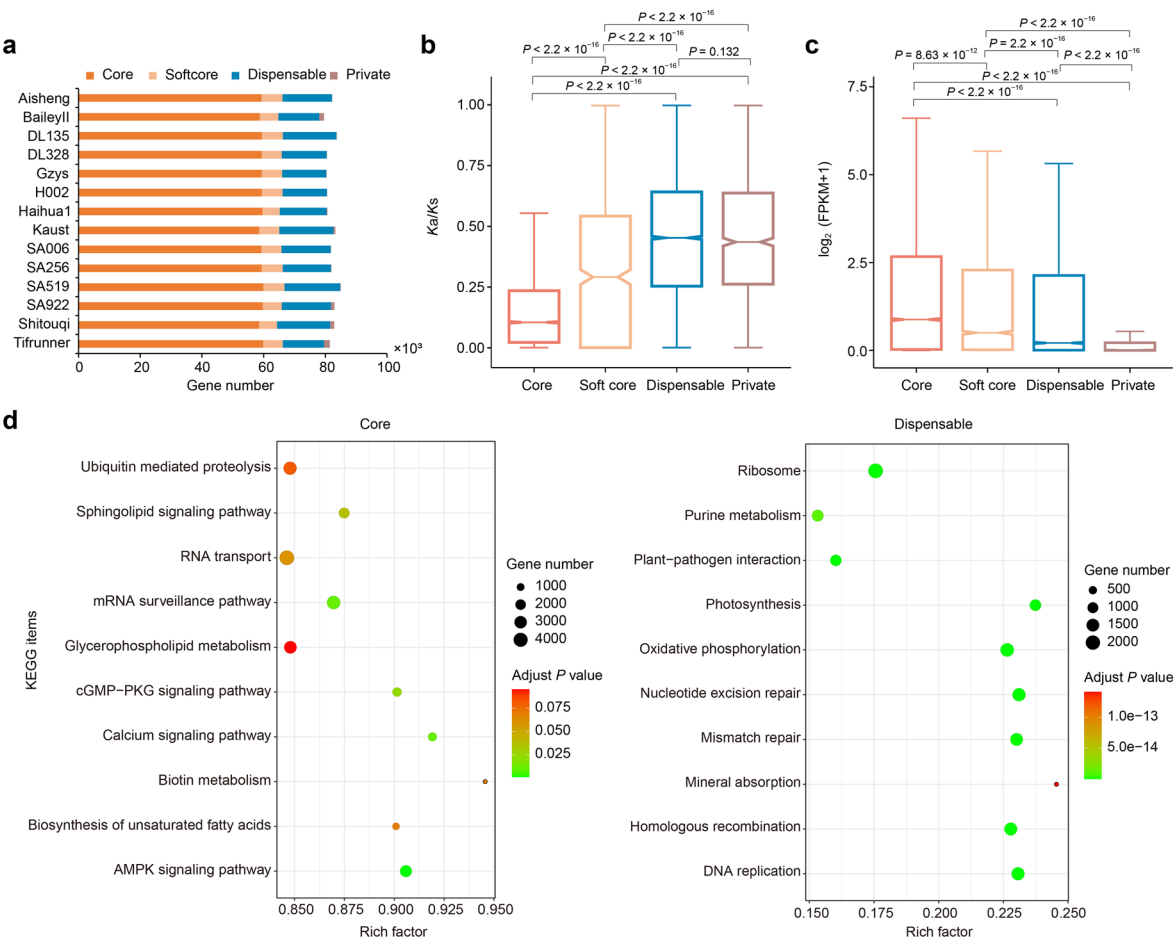
**Extended data** is available for this paper at <https://doi.org/10.1038/s41588-026-02765-x>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41588-026-02765-x>.

**Correspondence and requests for materials** should be addressed to Guowei Li, Xinyou Zhang, Rajeev K. Varshney or Shubo Wan.

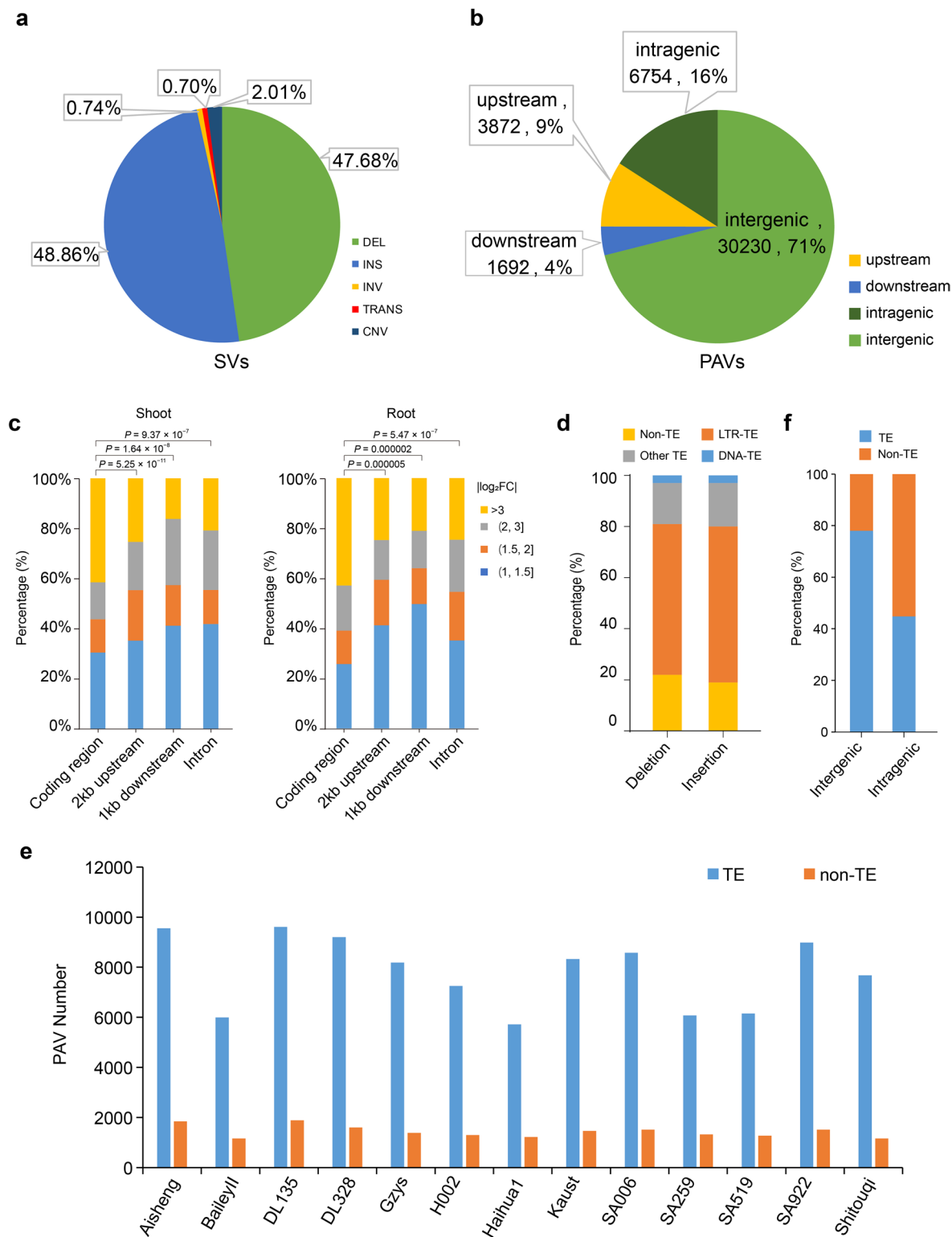
**Peer review information** *Nature Genetics* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



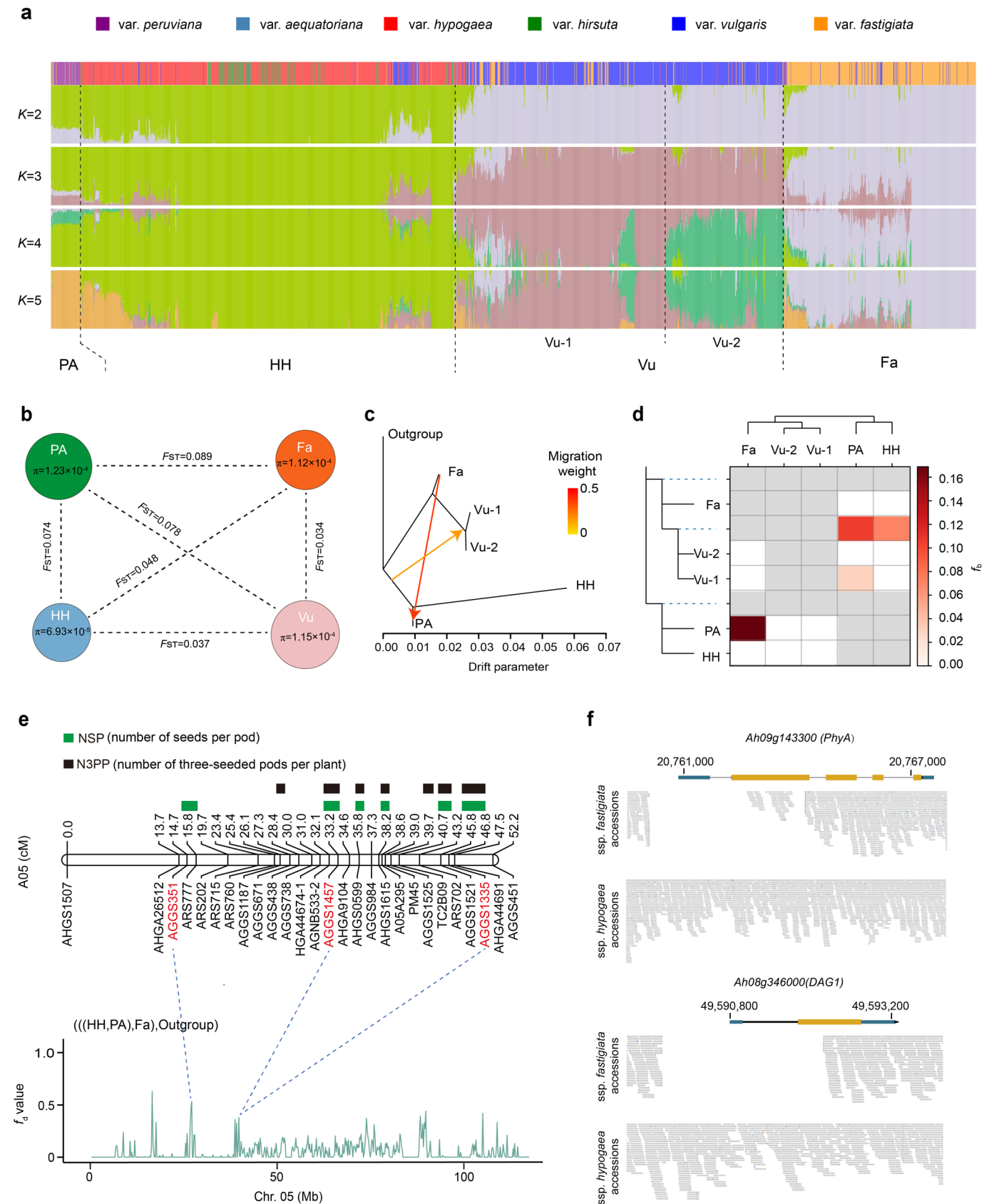
**Extended Data Fig. 1 | Pan-genome analysis of 14 genomes. a**, Number of core, softcore, dispensable and private genes across the 14 genomes. **b**, Comparison of *Ka/Ks* values of core, softcore, dispensable, and private genes. **c**, Expressions of core, softcore, dispensable and private genes across 14 accessions. For **b** and **c**, the same gene sets were analyzed: core genes, *n* = 48,948; softcore genes, *n* = 6,427; dispensable genes, *n* = 31,541; and private genes, *n* = 1,817. Statistical

significance was assessed using a Kruskal–Wallis rank-sum test followed by Dunn’s multiple-comparisons test with Benjamini–Hochberg adjustment. Boxes show the median and interquartile range, and whiskers extend to 1.5 times the interquartile range. **d**, KEGG pathway enrichment analysis of core genes (left) and dispensable genes (right). *P* values were false discovery rate (FDR)-adjusted.



**Extended Data Fig. 2 | Impact of structural variants (SVs) and presence-absence variants (PAVs) on gene.** **a**, Percentage of different types of structural variants (SVs) identified across 14 genomes (deletions = 21,015, insertions = 21,534, copy number variations = 887, inversions = 328 and translocations = 309). **b**, Presence-absence variants (PAVs) are predominantly located in intergenic regions (71%), with 16% in intragenic regions, 9% upstream, and 4% downstream of genes. **c**, Fold change in gene expression between accessions with and

without PAVs in the coding region, intron, 2 kb upstream of the start codon, and 1 kb downstream of the stop codon of protein-coding genes in both root and shoot tissues. Statistical significance was assessed using a two-sided Pearson's chi-squared test. **d**, Proportion of TE-derived and non-TE PAVs in the peanut genomes. **e**, Percentage of TE and non-TE PAVs in intergenic and intragenic regions. **f**, Distribution of TE PAVs and non-TE PAVs across accessions.

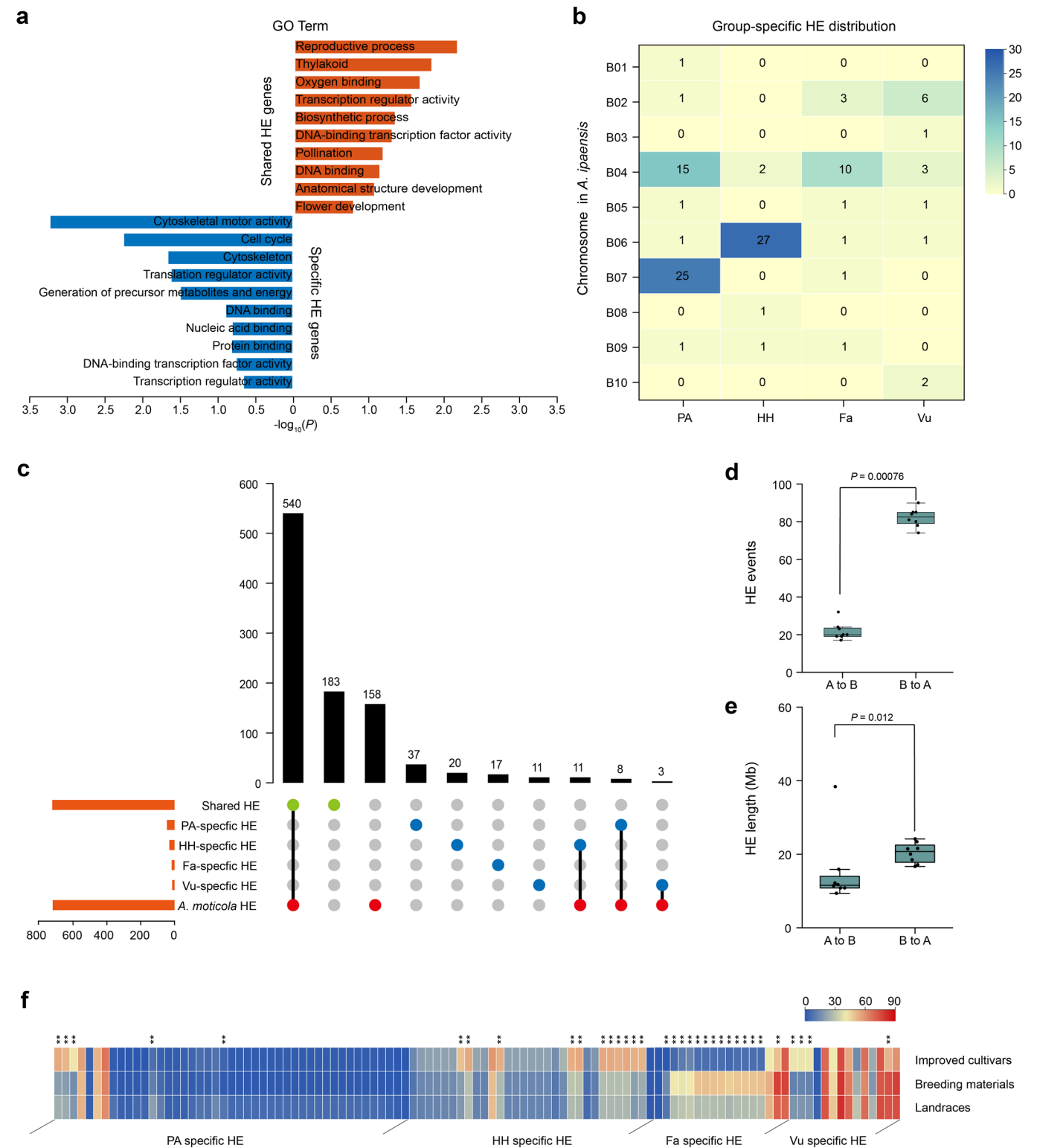


Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | Population structure, population divergence, gene flow, and structural variant (SV) number analysis across different groups.**

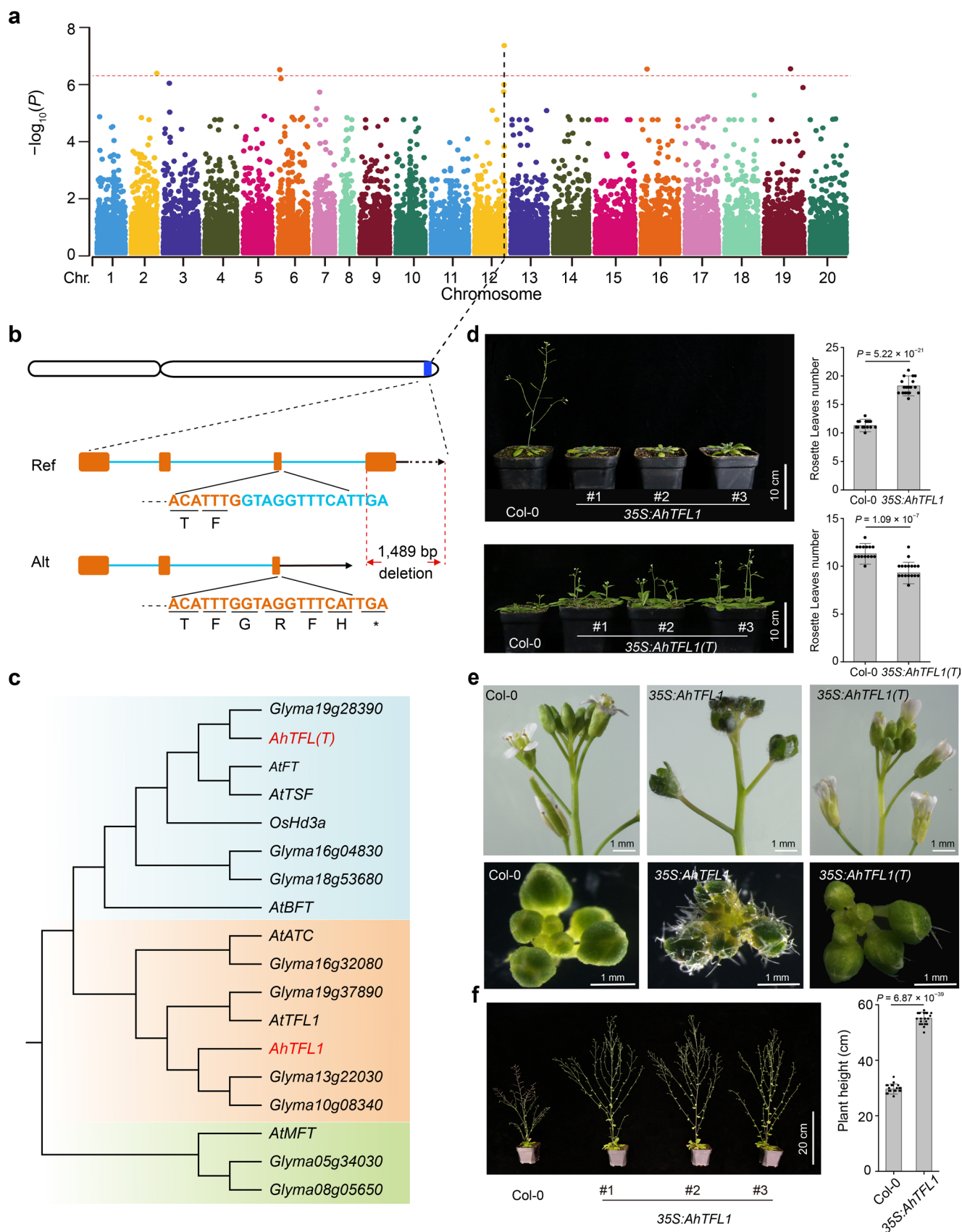
**a**, Population structure analysis using the STRUCTURE algorithm for  $K = 2-5$ . Each vertical bar represents a single accession, and the color proportions within the bar correspond to its genetic ancestry from different clusters. **b**, Population divergence ( $F_{ST}$ ) and nucleotide diversity ( $\pi$ ) among four groups (PA, HH, Vu, and Fa). **c**, Gene flow among different peanut populations. Branch lengths are scaled to genetic drift. Colored arrows indicate the direction of gene flow, with the color gradient representing the estimated migration weight. **d**, Signal of introgression estimated by the  $f$ -branch ( $f_b$ ) statistic. The heatmap displays  $f_b$  values calculated via Dsuite. Darker colors indicate stronger evidence of gene flow between

peanut populations (y-axis) and target taxa (x-axis). **e**, Integration of major QTL mapping for seed number per pod and Fa to PA introgression signals on Chr.05. Green boxes indicate QTLs associated with NSP (number of seeds per pod), and black boxes indicate QTLs associated with N3PP (number of three-seeded pods per plant). Previously reported QTLs were obtained from ref. 13 and re-plotted for comparison. Markers highlighted in red represent the linkage markers closest to the peak position of the major QTL. Dashed lines connect the major QTL region with overlapping introgressed intervals. **f**, Genomic variation of two subspecies differentiation genes, *Ah09g143300* (*PhyA*), and *Ah08g34600* (*DAG1*) in representative accessions.



**Extended Data Fig. 4 | Characteristics of shared and group-specific HE genes across different groups. a**, GO term enrichment analysis of shared and specific HE genes. The bar plot shows the top 10 enriched GO terms for shared (red) and specific (blue) HE genes. GO term enrichment was assessed using a one-sided Fisher's exact test.  $P$  values were not adjusted for multiple comparisons. Bar lengths represent  $-\log_{10}(P)$ . **b**, Heatmap of the distribution of group-specific HE genes across different cultivated peanut groups (PA, HH, Fa, Vu) on the *A. ipaensis* chromosomes. The color intensity represents the number of group-specific HEs identified. **c**, The number of shared and population-specific HE events across cultivated peanut populations and *Arachis monticola*. **d**, Comparison of HE event

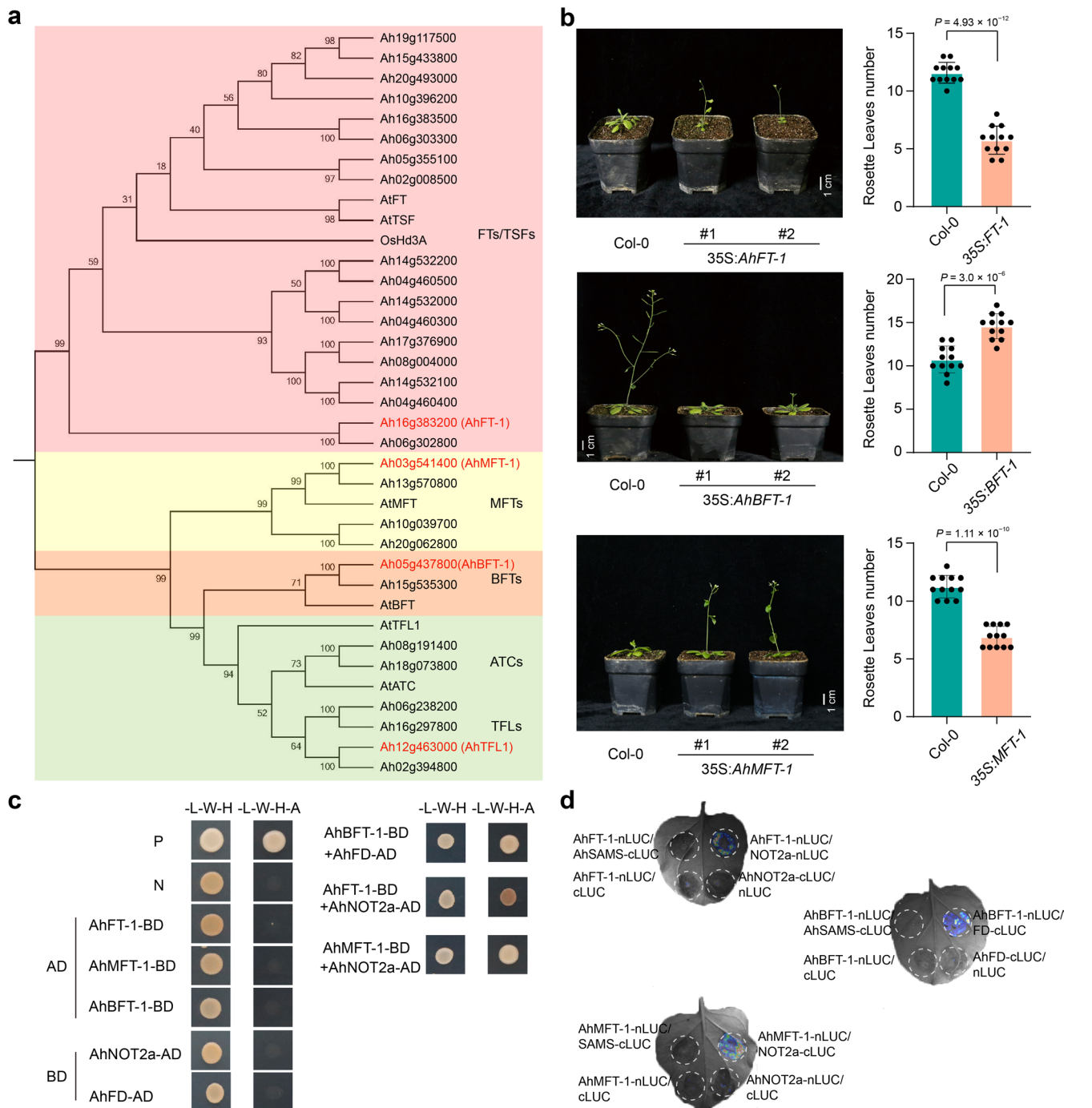
counts between A to B and B to A directions in *A. monticola*. **e**, Comparison of HE regions lengths (Mb) between A-to-B and B-to-A directions in *A. monticola*. For **d** and **e**,  $n = 8$  accessions per group. Boxes show the median and interquartile range, and whiskers extend to 1.5 times the interquartile range. **f**, Heatmap of the population frequencies of group-specific HE genes among 586 peanut accessions, including improved cultivars, breeding materials, and landraces. Frequencies are expressed as percentages within each population group. Differences in frequency among population groups were assessed using a two-sided Pearson's chi-squared test. Asterisks indicate significant differences based on unadjusted  $P$  values ( $P < 0.01$ ).



Extended Data Fig. 5 | See next page for caption.

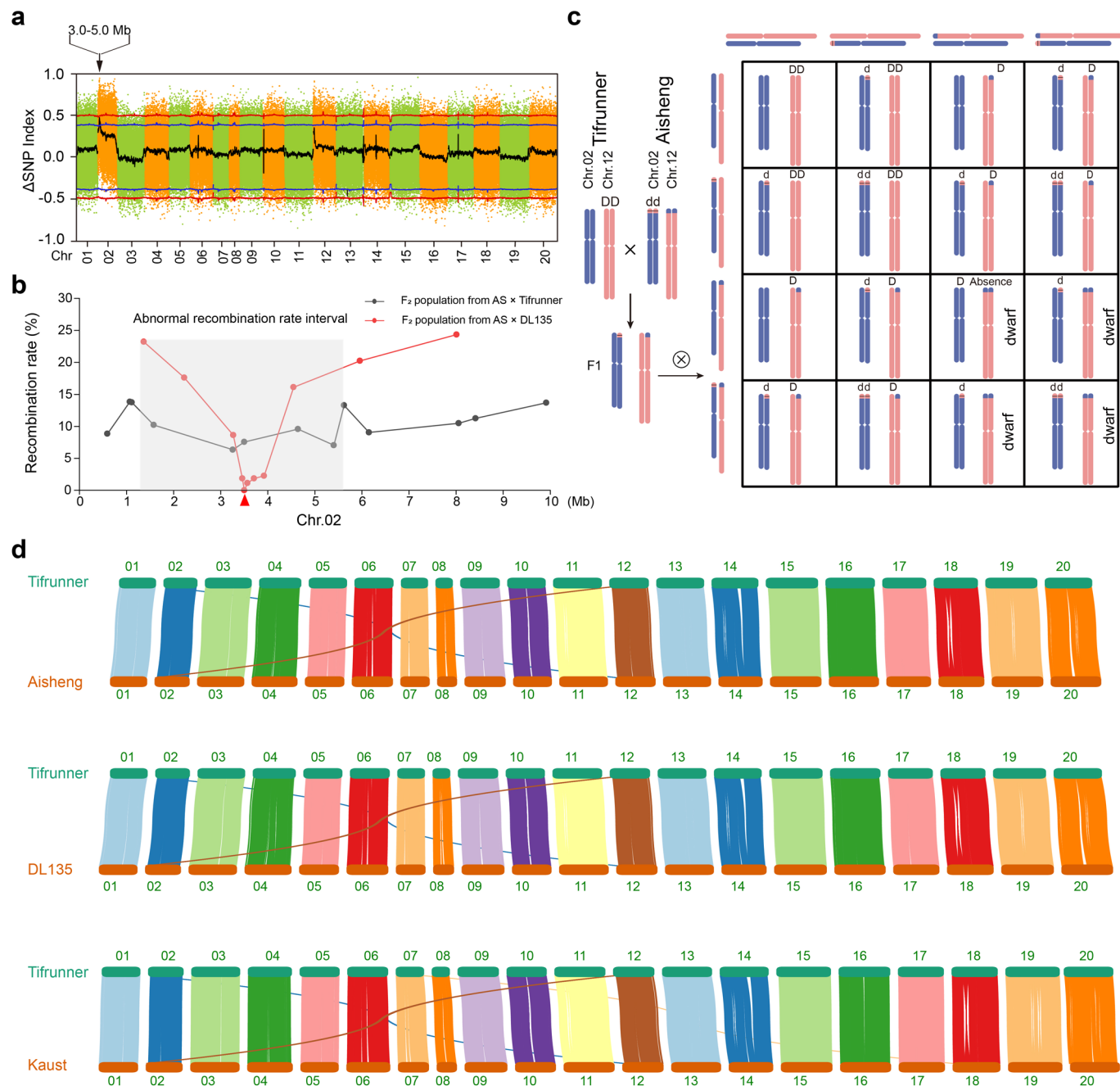
**Extended Data Fig. 5 | Functional analysis of *AhTFL1* and its role in flower development.** **a**, Manhattan plot of SV-GWAS. A significant association signal is identified on chr. 12, marked by the dotted red line. The horizontal red dashed line indicates the Bonferroni-adjusted genome-wide significance threshold ( $P = 5.53 \times 10^{-7}$ ). **b**, Diagram of the *AhTFL1* gene structure comparing Ref (Tifrunner) and Alt (Shitouqi). The Alt allele has a 1,489 bp deletion. The orange boxes represent exons, and the sequence alignment between the two alleles shows key differences in the nucleotide sequence. **c**, Phylogenetic tree of *TFL1* homologs. *AhTFL1* is closely related to *TFL1* genes, but *AhTFL1 (T)* is closer to *FT*

in other plant species. **d**, Flowering time and rosette leaf number in *Arabidopsis* lines overexpressing *AhTFL1* or *AhTFL1 (T)*. Scale bar, 10 cm. Data are presented as mean  $\pm$  s.d. ( $n = 24$  biologically independent samples). Statistical significance was assessed using a two-sided Student's *t*-test. **e**, Flower morphology of transgenic *Arabidopsis* lines overexpressing *AhTFL1* or *AhTFL1 (T)*. **f**, Plant height of *AhTFL1*-overexpressing lines and Col-0. Scale bar, 20 cm. Data are presented as mean  $\pm$  s.d. ( $n = 24$  biologically independent samples). Statistical significance was assessed using a two-sided Student's *t*-test.



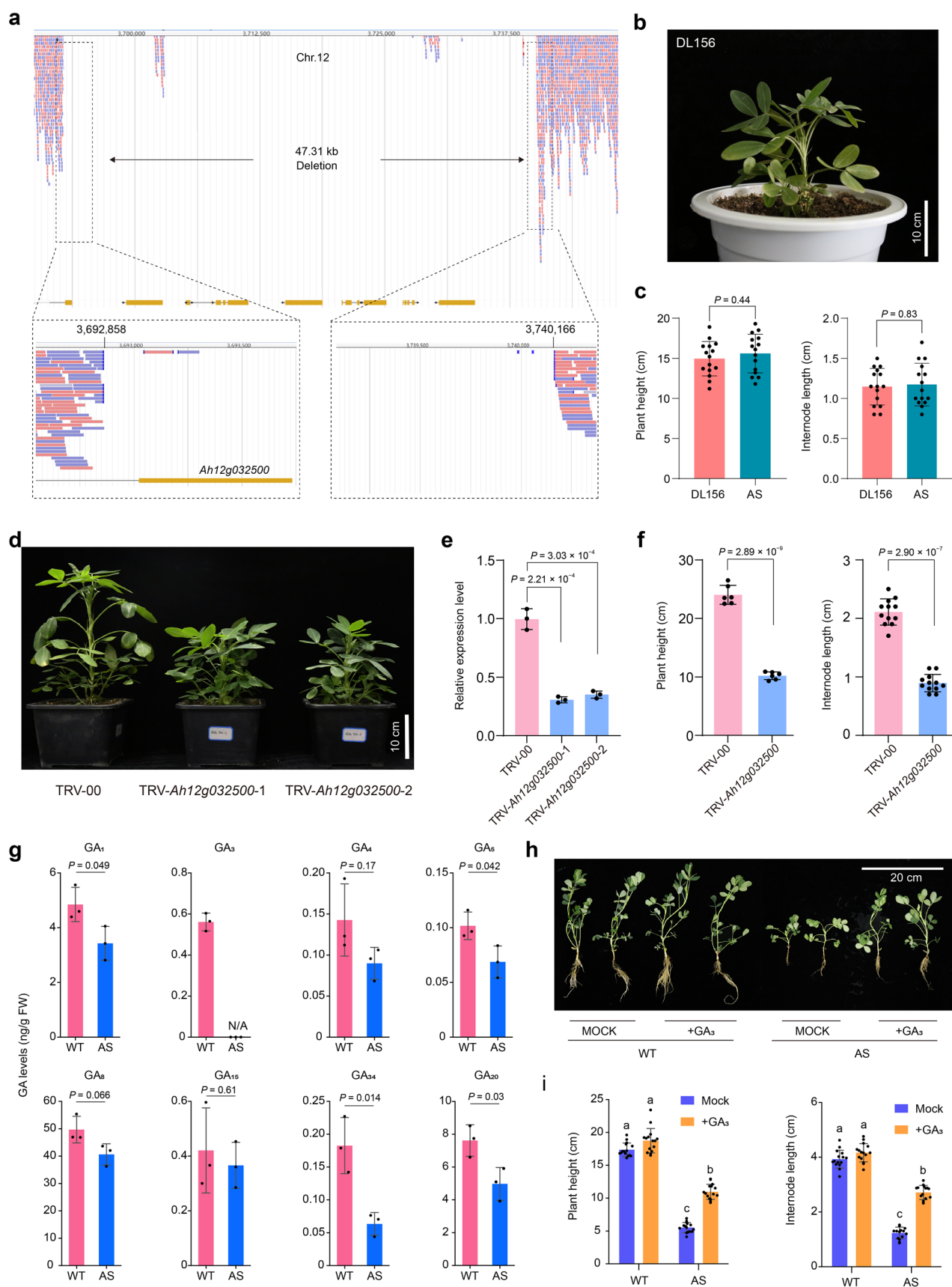
**Extended Data Fig. 6 | Phylogenetic analysis and functional characterization of *AhTFL*.** **a**, Phylogenetic tree of *FT/TFL* family genes in peanut and their homologues in *Arabidopsis*. Bootstrap support values are indicated on branches. **b**, Functional analysis of *35S:AhFT-1*, *35S:AhMFT-1* and *35S:AhBFT-1* transgenic *Arabidopsis* plants. Scale bar, 1 cm. Data are presented as mean  $\pm$  s.d. ( $n = 12$  biologically independent samples). Statistical significance was assessed using a two-sided Student's *t*-test. **c**, Yeast two-hybrid assay showing interactions of

*AhFT*, *AhMFT-1* and *AhBFT* proteins with other flowering regulators in peanuts. Yeast was grown on SD/-Leu/-Trp/-His (-L-W-H) and SD/-Leu/-Trp/-His/-Ade (-L-W-H-A) media to assess protein-protein interactions. **d**, The split-luciferase complementation (Split-LUC) imaging assay in *Nicotiana benthamiana* leaf cells confirms interactions between *AhFT*, *AhMFT-1* and *AhBFT* proteins with other flowering regulators. In **c** and **d**, representative images from three independent experiments with similar results are shown.



**Extended Data Fig. 7 | Abnormal genetic recombination in the hybrid population of Tifrunner and AS.** **a**, ΔSNP index analysis on F<sub>2</sub> population using AS and Tifrunner as parents. The ΔSNP index was calculated to identify genomic regions associated with dwarf phenotypes. A significant peak is observed on chr. 02 in the 3–5 Mb region. **b**, Abnormal and normal recombination rate from 0 to 6.0 Mb on chr. 02. **c**, Schematic representation of the abnormal hybrid

recombination between Tifrunner and AS, due to the translocation at dwarf gene region. **d**, Syntenic analysis of chr. 1–20 between Tifrunner and three representative accessions (AS, DL135, and Kaust). The most prominent balanced reciprocal translocation between chr. 02 and 12 was observed in AS, DL135 and Kaust compared to Tifrunner.



Extended Data Fig. 8 | See next page for caption.

**Extended Data Fig. 8 | Genotypic and phenotypic analysis of DL156, and GA-related analysis in AS plants.** **a**, Verification of the 47.31 kb deletion in DL156 using resequencing data. **b**, Morphology of DL156 at 50 days post-germination. Scale bar, 10 cm. **c**, Comparison of plant height and internode length between AS and DL156 at 90 days post-germination. Data are presented as mean  $\pm$  s.d. ( $n = 15$  biologically independent plants per variety). **d**, Phenotypic comparison of TRV-00 (empty vector control) and two independent TRV-*Ah12g032500* silencing lines. Scale bar, 10 cm. **e**, qRT-PCR analysis showing the relative expression levels of *Ah12g032500* in the TRV-00 and two silenced lines. Data are presented as means  $\pm$  s.d. from three biological replicates. **f**, Comparison of plant height and internode length between TRV-00 and TRV-*Ah12g032500* silencing

lines. Data are presented as mean  $\pm$  s.d. ( $n = 6$  biologically independent samples for plant height;  $n = 12$  biologically independent samples for internode length). **g**, Gibberellin contents in WT and AS seedlings. N/A, not detected. Data are presented as mean  $\pm$  s.d. ( $n = 3$  biologically independent samples). **h**, Growth of WT and AS seedlings in response to exogenous GA<sub>3</sub>. Three-week-old hydroponically cultured seedlings were treated with 10 mg GA<sub>3</sub> for 7 d. Scale bar, 20 cm. **i**, Plant height and internode length of WT and AS seedlings after GA<sub>3</sub> treatment. Data are presented as mean  $\pm$  s.d. ( $n = 15$  biologically independent samples). For **c**, **e**, **f** and **g**, statistical significance was determined by two-sided Student's *t*-test. For **i**, different letters indicate significant differences determined by one-way ANOVA followed by Tukey's test;  $P < 0.05$ .

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

|                                     |                                                                                                                                                                                                                                                                                                |
|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| n/a                                 | Confirmed                                                                                                                                                                                                                                                                                      |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( <i>n</i> ) for each experimental group/condition, given as a discrete number and unit of measurement                                                                                                                               |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly                                                                                                                                    |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>                                                               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested                                                                                                                                                                                                                                |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons                                                                                                                                                   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted<br><i>Give <i>P</i> values as exact values whenever suitable.</i>              |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings                                                                                                                                                                      |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes                                                                                                                                                |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i> ), indicating how they were calculated                                                                                                                                                          |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

|                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data collection | Data were sequenced from PacBio, Nanopore, Illumina and BGI platform.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Data analysis   | We used publicly available and appropriately cited software as described. No commercial software or code was used in this study. Software are listed as follows: GUPPY software (v0.3.0); Juicer (v1.6); Hifiasm (v0.16.1-r375); unimap software (v0.1-r46-dirty); LR_Gapcloser; Extensive de-novo TE Annotator (EDTA) (v2.1.3); RepeatMasker (v4.07); MAKER2 pipeline (v2.31.11); AUGUSTUS (v3.4.0); HISAT2 (v2.1.0); StringTie2 (v2.1.7);featureCounts (v2.0.6); Program to Assemble Spliced Alignments (PASA) (v2.4.1); EvidenceModeler (EVM) (v1.1.1); EggNOG-mapper (v1.0.3); DIAMOND (v0.9.24.125); BWA-MEM (v0.7.17); BUSCO (v2.0); LTR_retriever (v.3.0.1); Merqury (v.1.3); HiC-Pro (v2.9.0); HiCPlotter (v0.8.1); OrthoFinder (v2.5.5); KaKs_Calculator (v2.0); MUMmer toolkit (v4.0.0rc1); SyRI (v1.5.4); VCFtools (v0.1.17); SURVIVOR (v1.0.7); plotsr (v1.1.1); BEDTools (v2.31.0); FastTree (v2.1.11);MCScanX; GMAP; vg toolkit (v1.49.0); PLINK (v1.90p); PHYLP (v3.69); FigTree (v1.4.3); ADMIXTURE (v1.3.0); fastsimcoal2 (v2.6.0.3-14.10.17); BWA-MEM (v0.7.17); samtools coverage (v1.11); fastp (v0.23.0); Genome Analysis Toolkit (GATK) HaplotypeCaller (v4.4.0.0); SnpEff (v5.0e); rMVP software package(v1.0.8); TreeMix (v1.13); vcf2phylip.py ( <a href="https://github.com/edgardomortiz/vcf2phylip">https://github.com/edgardomortiz/vcf2phylip</a> ); QTL-seq ( <a href="https://github.com/omicsclass/QTL-seq">https://github.com/omicsclass/QTL-seq</a> ); Dsuite software. ABBABABAwindows.py (no versions, <a href="https://github.com/simonhmartin/genomics_general">https://github.com/simonhmartin/genomics_general</a> ); Primer3.0 ( <a href="https://primer3.ut.ee/">https://primer3.ut.ee/</a> ). The codes used in this project were available at GitHub <a href="https://github.com/lrcchina/Peanut-pangenome">https://github.com/lrcchina/Peanut-pangenome</a> and Zenodo ( <a href="https://doi.org/10.5281/zenodo.13906381">https://doi.org/10.5281/zenodo.13906381</a> ). |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The genome sequencing and assembly data of 10 accessions of cultivated peanut have been deposited at National Center for Biotechnology Information under the BioProject PRJNA1168101, PRJNA1170232, PRJNA1169541, PRJNA1170953. Transcriptome data from different tissues of 10 accessions of cultivated peanut can be found under the BioProject PRJNA1167437. The whole-genome sequencing of 2,320 accessions has been deposited at NCBI under the BioProject PRJNA1163959 and PRJNA1071845. RNA-seq data from 22 different tissues as well as various abiotic stresses in cultivated peanut were downloaded from NCBI under BioProject PRJNA291488 and BioProject PRJNA553073. Source data are provided with this paper.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

|                                                                    |                 |
|--------------------------------------------------------------------|-----------------|
| Reporting on sex and gender                                        | Not applicable. |
| Reporting on race, ethnicity, or other socially relevant groupings | Not applicable. |
| Population characteristics                                         | Not applicable. |
| Recruitment                                                        | Not applicable. |
| Ethics oversight                                                   | Not applicable. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

|                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sample size     | No statistical method was used to predetermine sample size. 14 representative peanut accessions were used for pan-genome construction. These accessions were selected based on their phylogenetic relationship and represented genetic diversity in the peanut germplasm. The sample size of 2,320 peanut accessions for resequencing were chosen from 87 countries and regions, including 1,500 core germplasm accessions from ICRISAT, 767 from the USDA Peanut Germplasm Resource Collection, and others, representing six botanical varieties of cultivated peanut. |
| Data exclusions | No data were excluded from the analysis.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| Replication     | Independent replication was performed where appropriate for each experiment. The number of independent experiments or biological replicates is specified in the corresponding figure legends and Methods.                                                                                                                                                                                                                                                                                                                                                               |
| Randomization   | For trait evaluation, we randomly sampled six plants out of 50 plants per accession in the field. For DNA extractions, we randomly sampled 1 plant out of 50 plants in the field.                                                                                                                                                                                                                                                                                                                                                                                       |
| Blinding        | Blinding was not performed because the study involved predefined plant materials, breeding populations and computational analyses, for which blinding was not applicable.                                                                                                                                                                                                                                                                                                                                                                                               |

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a                                 | Involved in the study                                  |
|-------------------------------------|--------------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Plants             |

## Methods

| n/a                                 | Involved in the study                           |
|-------------------------------------|-------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

## Dual use research of concern

Policy information about [dual use research of concern](#)

### Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

| No                                  | Yes                                                 |
|-------------------------------------|-----------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Public health              |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> National security          |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Crops and/or livestock     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Ecosystems                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Any other significant area |

### Experiments of concern

Does the work involve any of these experiments of concern:

| No                                  | Yes                                                                                                  |
|-------------------------------------|------------------------------------------------------------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Demonstrate how to render a vaccine ineffective                             |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Confer resistance to therapeutically useful antibiotics or antiviral agents |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enhance the virulence of a pathogen or render a nonpathogen virulent        |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Increase transmissibility of a pathogen                                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Alter the host range of a pathogen                                          |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enable evasion of diagnostic/detection modalities                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enable the weaponization of a biological agent or toxin                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Any other potentially harmful combination of experiments and agents         |

## Plants

|                       |                                                                                                                                                                                                                                                                                                       |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Seed stocks           | All 2,320 accessions used stocked in the open germplasm stocks of the International Crops Research Institute for the Semi-Arid Tropics in India, the United States Department of Agriculture, and Shandong Academy of Agricultural Sciences in China.                                                 |
| Novel plant genotypes | Transgenic plants were then created using the Agrobacterium-mediated floral-dip method. The virus-induced gene silencing (VIGS) system based on the tobacco rattle virus (TRV) methodology. Two or three homozygous lines of transgenic plants at T3 generation were used for analysis in this study. |
| Authentication        | The transgenic Arabidopsis plants were authenticated by PCR amplification and DNA sequencing methods. The efficiency of virus-induced gene silencing (VIGS) was verified by qRT-PCR analysis of target gene expression levels.                                                                        |