Lineage-specific evolution of regulatory landscapes in polyploid plant and its diploid progenitors

Xiang Li[1], Xuan Zhang[1], Ziliang Luo[1], Hao Zhang[1], John Pablo Mendieta[1], Robert J. Schmitz[1]

[1]Department of Genetics, University of Georgia, Athens, GA, 30602, USA.

Email: schmitz@uga.edu

Abstract

Cis-regulatory elements are specific DNA sequences that control gene expression in a spatiotemporal manner, and variation within these elements represents a major source of phenotypic diversity and evolutionary innovation. Nevertheless, how regulatory elements evolve and shape gene expression remains poorly understood, particularly in plants. The well-resolved phylogeny of allopolyploid peanut (*Arachis hypogaea*) and its diploid progenitors, *A. duranensis* and *A. ipaensis*, provides an ideal system to investigate the regulatory evolution at a lineage-specific level. By integrating comparative analyses of sequence similarity, chromatin accessibility, histone modifications, conserved noncoding sequences (CNSs), and gene expression, we reconstructed the evolutionary trajectories of Accessible Chromatin Regions (ACRs), where regulatory elements typically reside, and revealed their distinct contributions to homeolog expression bias, unequal expressions between homeologs, in a polyploid genome. Most ACRs exhibited high sequence similarity, comparable chromatin accessibility, and conserved states for H3K4me3, H3K56ac, and H3K36me3, indicating regulatory stability after hybridization and polyploidization. However, a subset of novel ACRs emerged *de novo* or through sequence mutations in preexisting ACRs, arising at different rates and evolutionary stages. Notably, even highly sequence-conserved ACRs exhibited substantial variation in chromatin accessibility variation, consistent with CNS composition differences and minor sequence variation, although causal relationships remain to be demonstrated. Our analyses further revealed a complex spectrum of CNS dynamics across the diploid-polyploid framework. Overall, our study provides empirical insights into the fine-scale evolution of plant regulatory landscapes and complements previous large-scale comparisons across distant lineages.

Introduction

Cis-regulatory elements, specific DNA sequences bound by transcription factors, regulate gene expression in a spatiotemporal manner (1–4). Genetic variation within these elements represents an important source of phenotypic diversity and evolutionary innovation (5, 6). Nevertheless, the origin and the extent of sequence divergence within these elements and how such divergence shapes regulatory activity and gene expression remain poorly understood, especially in plants.

Unlike the relatively conserved and stable mammalian genomes, plant genomes are highly dynamic, exhibiting extensive variation in size, structure, and compaction (7, 8). Frequent macro- and micro- genomic rearrangements, whole genome duplication, gene duplication and unequal gene retention have all shaped plant genome architecture. These processes likely exert a profound impact on the emergence and diversification of regulatory sequences (9–11). Comparative analyses across distantly related plants have revealed a high rate of turnover in regulatory regions, underscoring the fast-evolving nature of plant *cis*-regulatory landscapes and

the difficulty of tracing their evolutionary trajectories (12–14). Yet, deeply conserved non-coding regions (CNSs) spanning more than 280 plant species and 300 million years of diversification have also been identified (10). To reconcile this seemingly contradictory observation between, rapid turnover versus deep conservation, and to complement broad-scale comparative studies, analyses among closely related species provides a critical and trackable framework to uncover how *cis*-regulatory sequences diversify, acquire new functions, and drive regulatory innovation over short evolutionary timescales.

Polyploid plants that have undergone whole-genome duplication provide a natural context to contrast the evolutionary fates of duplicated regulatory elements. During polyploidization, regulatory elements are copied alongside genes and may experience relaxed selective constraints like those acting on duplicated coding sequences (15), thereby enabling divergence in both sequence and function (5). Recent studies in soybean, cotton, and strawberry have illustrated that variation in *cis*-regulatory regions contributes to differential expression among duplicated genes (16–22). However, most insights come from species with ancient duplication events or from analyses lacking direct comparisons with their progenitors. Incorporating progenitor species with polyploid species provides a unique opportunity to reconstruct how regulatory sequences originate and diversify.

This framework is particularly well suited for recently formed allopolyploids such as cultivated peanut (*Arachis hypogaea*), an allotetraploid crop (AABB, 2n = 4x =40) derived from hybridization between two diploid progenitors, *A. duranensis* (progenitor A) and *A. ipaensis* (progenitor B), followed by whole-genome duplication (23–29). Two diploid progenitors diverged ~2 million years ago (Mya), whereas divergence between each A. *hypogaea* subgenome and its corresponding progenitor occurred ~0.2 Mya (28). This short evolutionary timespan has resulted in two subgenomes being clearly distinguishable (25), and likely insufficient time for extensive *cis*-regulatory divergence between derived and ancestral states. Peanut, therefore, represents an exceptional model to examine the early stages of *cis*-regulatory evolution and to trace how regulatory elements evolve through genome hybridization and duplication.

To investigate sequence variation in *cis*-regulatory elements and its relationship to regulatory activity and gene expression divergence, we analyzed three peanut species: *A. hypogaea*, *A. duranensis* and *A. ipaensis*, using both leaf and root tissues. We integrated Assay for Transposase Accessible Chromatin Sequencing (ATAC-seq) to map genome-wide chromatin accessibility, Chromatin Immunoprecipitation Sequencing (ChIP-seq) to profile the histone modifications, and RNA-seq to quantify transcript abundance. We identified Accessible Chromatin Regions (ACRs) and classified them into distinct evolutionary categories based on sequence conservation. Comparative analyses of sequence variation, chromatin accessibility states, histone modification patterns and CNSs dynamics, revealed distinct evolutionary and functional signatures among these classes. Together, these results illustrate the dynamic regulatory landscapes of recently formed polyploids, where ACRs of different evolutionary origins follow diverse trajectories that collectively shape regulatory balance and expression diversification within the polyploid genome.

Results

Mapping the accessible chromatin landscape in peanut

We generated 12 ATAC-seq libraries, comprising two biological replicates of leaf and root tissues from each of the three peanut species: *A. hypogaea*, *A. duranensis* and *A. ipaensis*. An average of 84,513,689 raw reads were obtained per library. After adapter trimming and quality filtering, uniquely mapped reads were used for peak calling, and regions meeting defined thresholds were designated as ACRs (see Methods). On average 61,539 peaks were identified per replicate (**Table S1**). The average proportion of reads around Transcription Start Site (TSS) and average Fraction of Reads in Peak (FRiP) score were 0.35 and 0.48, indicating the good quality of ATAC-seq data (**Table S1**).

Comparative characterization of ACRs across four peanut genomes

We merged ACRs identified from leaf and root tissues for each genome: subgenome A and subgenome B of *A. hypogaea* and the two diploid progenitors, *A. duranensis* and *A. ipaensis*, yielding 56,526, 56,847, 113,025, and 95,584 ACRs, respectively (Table S2). These merged ACRs were used as references for differential chromatin accessibility analysis between leaf and root tissues. Regions showing no significant difference in chromatin accessibility were classified as broad ACRs, whereas those with significantly higher chromatin accessibility in either tissue were designated as leaf-specific or root-specific (**Fig. 1A**).

Broad ACRs accounted for the majority of ACRs, with tissue-specific ACRs representing a smaller fraction (**Fig. 1B, Fig. S1** and **Table S2**), consistent with findings from previous studies (13, 14, 30–32). Based on proximity to genes, ACRs were further categorized as genic, proximal (≤2 kb from genes), or distal (>2 kb). Among tissue-specific ACRs, distal regions were most prevalent, whereas broad ACRs were enriched in genic and proximal regions (**Fig. 1C, Fig. S1** and **Table S2**). Although a greater number of ACRs were identified in diploid genomes, both diploids and two subgenomes exhibited comparable ACR patterns of tissue specificity and genomic distribution (**Fig. 1B-1C, Fig. S1** and **Table S2**).

Identification of ACRs with distinct evolutionary trajectories

The well-resolved phylogenetic relationships between the two subgenomes and their respective diploid progenitors provide a rigorous framework for tracing the evolutionary trajectories of *cis*-regulatory elements (**Fig. 2A**). To infer the sequence ancestry of ACRs in polyploid peanut, we performed BLAST searches of ACRs from each subgenome against the other subgenome and both diploid progenitor genomes. To preserve the local genomic context of each regulatory region, we implemented a synteny-based BLAST pipeline adapted from (13). In this framework, each query ACR was anchored by its flanking homoeologous genes, and BLAST searches were restricted to the corresponding syntenic intervals in the target genomes (**Fig. S2**). We focused on ACRs located within syntenic intervals conserved across all four genomes: the two subgenomes and two diploid genomes. Although restricting analyses to conserved syntenic intervals may miss non-collinear homologous ACRs and transposable elements (TEs)-mediated relocations (33, 34), it provides a robust baseline for tracking regulatory evolution. Based on sequence conservation profiles, ACRs were classified into five evolutionary categories reflecting their possible origins: 1) **subgenome-specific**, present only in one subgenome and absent from the other subgenome and both diploid genomes, suggesting *de novo* emergence; 2) **polyploid-**

134 **specific**, shared between subgenomes but not detected in diploids, indicating formation after
135 hybridization but prior to genome duplication; 3) **lineage-specific**, conserved only between
136 subgenome and its corresponding diploid progenitor, representing inheritance from the
137 ancestral genome; 4) **peanut-conserved**, retained across four genomes, indicating long-term
138 retention throughout the peanut lineage; 5) **potential peanut-conserved**, detected in three
139 genomes or shared between a subgenome and the non-corresponding diploid genome (**Fig. 2A**).
140
141 To evaluate the stability of these classifications, we applied two sequence similarity thresholds
142 (≥50% and ≥80%). Raising the threshold from 50% to 80% led to a marked reclassification of
143 ACRs, particularly with many shifting from peanut-conserved to lineage-specific categories (**Fig.**
144 **2B, Fig. S3** and **Table S3**). This shift indicates that sequence divergence has accumulated within
145 regions previously considered conserved, reflecting their recent evolutionary trajectories.
146 Because the higher threshold more accurately captures these recent differences, subsequent
147 analyses were based on ACR classifications defined by the 80% criterion. In both subgenomes,
148 most ACRs within conserved syntenic intervals were peanut-conserved (68%) (**Fig. 2B, Fig. S3**
149 and **Table S3**). When including potential peanut-conserved ACRs, this proportion increased to
150 approximately 80%, underscoring their overall high sequence conservation. Lineage-specific
151 accounted for ~12%, followed by subgenome-specific (~2-4%) and polyploid-specific ACRs (~2-
152 4%) (**Fig. 2B, Fig. S3** and **Table S3**). Notably, subgenome-specific ACRs were more abundant in
153 subgenome A (4%) than in subgenome B (2%) (**Fig. S3** and **Table S3**).
154
155 Tissue specificity patterns were largely consistent across categories, with most ACRs exhibiting
156 broad chromatin accessibility (**Fig. S4** and **Table S4**). Genomic distributions were also
157 comparable, although peanut-conserved ACRs, potential peanut-conserved ACRs and polyploid-
158 specific ACRs had slightly higher proportions of genic ACRs (**Fig. S4** and **Table S4**).
159
160 **TEs disproportionately shape ACRs with different evolutionary trajectories**
161 Repeat sequences are abundant in the peanut genome, with retroelements and DNA
162 transposons comprising the majority (29). TEs are well known to drive the emergence of novel
163 *cis*-regulatory elements by introducing regulatory sequences into new genomic contexts (5, 12,
164 33–37). In our analysis, TE associations were more frequent among subgenome-specific ACRs
165 (**Fig. 2C**, **Fig. S5** and **Table S4**). For instance, 85.8% of subgenome-specific ACRs in subgenome B
166 overlapped with repeat elements by at least 20 bp, and 20.7 % were entirely embedded within
167 repeats (**Fig. 2C**, **Fig. S5** and **Table S4**), underscoring the strong contribution of TEs to regulatory
168 novelty. Moreover, TE-associated ACRs were consistently more abundant in subgenome B than
169 in subgenome A (**Fig. S5** and **Table S4**), suggesting that TE insertions may have exerted
170 asymmetric influences on the evolution of regulatory landscapes between the two subgenomes.
171
172 **Genetic variation patterns among ACRs with distinct evolutionary histories**
173 To evaluate sequence variation among ACRs with different evolutionary trajectories, we
174 quantified two metrics: the BLASTed ratio, defined as the length of the query ACR divided by the
175 length of its BLASTed sequence in the corresponding reference region, and the mismatch rate,
176 defined as the number of mismatches per BLASTed sequence. Subgenome-specific ACRs were
177 excluded from this analysis because no corresponding sequences exist in the reference regions.
178 Across the remaining categories, most ACRs exhibited high BLASTed ratios (~1) and low
179 mismatch rates (~0.01-0.02) (**Fig. 2D, Fig. S6** and **Table S4**). These values contrast sharply with

180 those reported for conserved ACRs in soybean, where the median BLASTed ratio of recently
181 duplicated ACRs was ~0.75 and the median mismatch rates was ~0.05 (22). This comparison
182 underscores the elevated regulatory sequence conservation in this recently formed polyploids
183 and highlights the influence of divergence time on regulatory sequence decay. Interestingly,
184 lineage-specific ACRs exhibited the highest BLASTed ratios and lowest mismatch rates, whereas
185 polyploid-specific ACRs displayed the lowest BLASTed ratios and highest mismatch rates (**Fig.**
186 **2D, Fig. S6** and **Table S4**). This trend is somewhat counterintuitive, as lineage-specific ACRs likely
187 emerged earlier and might therefore be expected to accumulate more divergence. Their
188 elevated conservation may instead reflect stronger selective pressure associated with their
189 single-copy status in the polyploid genome. In contrast, the duplicated nature of peanut-
190 conserved and polyploid-specific ACRs may lead to relaxed selective constraints, facilitating the
191 accumulation of mutations.
192
193 **Histone modification dynamics of sequence-conserved ACRs**
194 To characterize the histone modification landscape associated with ACRs, we performed ChIP-
195 seq for three peanut species in both leaf and root tissues. Chromatin-state conservation was
196 inferred from three active histone marks: H3K4me3, H3K56ac, and H3K36me3. We did not
197 profile repressive histone marks or DNA methylation, which may exhibit distinct regulatory
198 dynamics (38, 39). Genome-wide patterns of histone marks were consistent with previous
199 reports (12, 40–43), with H3K4me3 and H3K56ac enriched around transcription start sites, and
200 H3K36me3 enriched within gene bodies (**Fig. S7**). We then intersected ACR coordinates with
201 peaks of individual histone marks and classified ACRs into three states: transcribed, Kac-
202 enriched, and unmodified (**Fig. 2E**), based on the combinatorial patterns of the three marks,
203 following the clustering criteria described in (12).
204 Comparison of histone states across peanut-conserved ACRs and their corresponding BLASTed
205 regions revealed that most ACRs (~85%) retained identical histone states between the two
206 subgenomes, and the majority also shared the same states as both diploid progenitors (~60%)
207 (**Fig. 2E** and **Table S5**). A small subset (~6%) displayed conserved states within subgenomes and
208 within progenitors but differed between subgenomes and progenitors (**Fig. 2E** and **Table S5**).
209 The remaining ~15% exhibited divergent histone states between subgenomes; in most of these
210 cases, one or both subgenomes matched the histone state of the corresponding progenitor,
211 highlighting strong parental-legacy effects, whereas only ~2% showed entirely novel histone
212 states absent from either progenitor (**Fig. 2E** and **Table S5**). Lineage-specific and polyploid-
213 specific ACRs also displayed high histone state similarity, ranging from 70% to 86%, either
214 between a subgenome and its corresponding progenitor or between the two subgenomes
215 (**Table S5**).
216
217 **Varied chromatin accessibility bias among sequence-conserved ACRs**
218 To further investigate the regulatory dynamics of sequence-conserved ACRs, we quantified
219 chromatin accessibility by counting reads overlapping each conserved ACR and its corresponding
220 BLASTed sequences. Peanut-conserved ACRs, those with highly conserved sequences across all
221 four genomes, provide a framework for examining chromatin accessibility dynamics in both
222 ancestral and derived states. Principal component analysis (PCA) of the normalized chromatin
223 accessibility counts revealed a strong parental-legacy effect on chromatin accessibility variation.
224 Principal Component 1 (PC1), explaining ~36.3% of the total variance, separated the two

225 lineages, subgenome A clustered with *A. duranensis*, and subgenome B with *A. ipaensis*. PC2,
226 accounting for ~27.5% of the variance, distinguished tissue types (**Fig. 3A** and **Fig. S8**).
227 Following the conceptual framework of homoeolog expression bias, in which homoeologous
228 genes exhibit differential expression (44–48), we examined chromatin accessibility bias in
229 peanut-conserved ACRs to determine whether chromatin accessibility differed between
230 subgenomes, whether the bias favored subgenome A or subgenome B, and whether such bias
231 reflected parental legacy or regulatory novelties that emerged following genome hybridization
232 and duplication. Accordingly, each ACR was classified into three major groups: (i) **parental**
233 **condition**, no-bias across all four genomes or a bias consistent with the progenitors; (ii) **no bias**
234 **in progeny**, a bias present in progenitors but absent in the polyploid; and (iii) **novel bias in**
235 **progeny**, bias distinct from the parental condition (**Fig. 3B**). Within each category, ACRs were
236 further designated as subgenome A- or subgenome B- dominant (or progenitor A- or progenitor
237 B- dominant) (**Fig. 3B**), depending on which genome exhibited significantly higher chromatin
238 accessibility (See Methods).
239
240 Distributions of ACRs with different chromatin accessibility bias patterns were largely similar
241 between leaf and root tissues, except that a greater proportion exhibited a parental no bias
242 pattern in root tissue (**Fig. 3C** and **Table S6**). Across tissues, most ACRs (~70–80%) showed no
243 bias between the two subgenomes, with the majority (~45–63%) also lacking bias between the
244 diploid progenitors (parental no bias), and the remainder (~13–23%) showing bias between the
245 progenitors (no bias progenitor A- or progenitor B- dominant) (**Fig. 3C** and **Table S6**).
246 Approximately 20% of ACRs exhibited significant accessibility bias between subgenomes,
247 primarily reflecting parental-legacy effects (parental subgenome A- or subgenome B- dominant
248 ~10%), followed by novel bias (novel bias subgenome A- or subgenome B- dominant ~6%) and
249 novel reverse bias (novel bias-r subgenome A- or subgenome B- dominant ~2%), in which the
250 direction of bias was opposite to that observed in the progenitors (**Fig. 3C** and **Table S6**).
251
252 Subgenome A- or subgenome B- dominant ACRs, regardless of whether they were classified as
253 parental, novel or novel reverse, were largely restricted to their respective subgenomes and
254 were typically not detected as ACRs in the other subgenome (**Fig. 3C** and **Table S6**). The
255 prevalence of these presence/absence differences in chromatin accessibility suggests that this
256 may represent a major form of regulatory asymmetry among sequence-conserved ACRs
257 between subgenomes. To investigate this, we overlapped the BLASTed coordinates of ACRs
258 identified from one subgenome (e.g., subgenome B) in the other subgenome (e.g., subgenome
259 A) with the coordinates of experimentally identified ACRs in the other subgenome (e.g.,
260 subgenome A). Subgenome A- or subgenome B- dominant ACRs exhibited markedly lower
261 overlap, whereas ACRs with other chromatin accessibility bias patterns showed much higher
262 overlap (**Fig. 3D, Fig. S9,** and **Table S7**). Furthermore, normalized chromatin accessibility counts
263 for these dominant ACRs were substantially higher in their own subgenome compared to their
264 BLASTed counterparts in the other subgenome (**Table S8**), reinforcing that differential
265 chromatin accessibility between subgenomes often manifests as presence-absence variation
266 rather than as quantitative shifts.
267
268 We next examined sequence variation among ACRs exhibiting different chromatin accessibility
269 bias patterns. Although no strong overall pattern was observed, parental no bias ACRs
270 consistently had the lowest mismatch rates and relatively high BLASTed ratio compared to ACRs

with altered chromatin accessibility (**Table S9**), suggesting that sequence divergence contributes to chromatin accessibility variation.

Comparisons of lineage-specific ACRs between each subgenome and its corresponding diploid progenitor, as well as of polyploid-specific ACRs between subgenomes, also revealed that most did not exhibit significant differences in chromatin accessibility (**Table S10**), consistent with previous observations of histone modification states (**Fig. 2E** and **Table S5**).

Age-dependent conservation and loss of CNSs in peanut subgenomes and progenitors

Cis-regulatory elements are often highly conserved, reflecting selection to maintain essential regulatory function (10, 49–53). To gain further insights into their evolutionary dynamics, we examined CNSs in the context of polyploid peanut and its corresponding diploid progenitors, providing complementary perspective to studies that have largely overlooked lineage-specific regulatory evolution (10, 36, 54, 55).

Consistent with previous findings (10), ancient CNSs, such as those conserved at the Tracheophyte (vascular plants) or Angiosperm level, were less numerous but exhibited higher copy numbers and greater retention than lineage-specific CNSs, such as family-level CNSs. For example, 71% of CNSs were conserved at the Fabaceae level, whereas only ~2% were conserved at the Angiosperm level (**Fig. 4A** and **Table S11**); ~75% of Tracheophyte-level CNSs had at least two copies, compared with 25% of Fabaceae-level CNSs (**Fig. 4B** and **Table S12**); and 55% of Tracheophyte-level CNSs were retained across all three peanut species, versus 23.9% of Fabaceae-level CNSs (**Fig. 4C** and **Table S13**). In contrast to the previously proposed view that one paralog largely retains ancestral CNSs while the other undergoes rapid turnover (10), we found that ~42% of Tracheophyte- and Angiosperm-level CNSs were preserved across both subgenomes and their diploid progenitors (**Fig. 4C** and **Table S13**). This pattern may partly reflect the relatively recent origin of the allotetraploid, leaving insufficient time for ancient CNSs to diverge.

Beyond these general trends, we observed a more complex spectrum of CNS dynamics across the diploid-polyploid framework. CNSs could be categorized as: (i) conserved across all four genomes, (ii) conserved across three genomes, (iii) conserved exclusively in progenitors, (iv) conserved exclusively in polyploid subgenomes, (v) conserved between a subgenome and its corresponding progenitor, (vi) conserved between a subgenome and the non-corresponding progenitor, and (vii) restricted to a single genome (**Fig. 4D-4E**).

Notably, the distribution of these patterns varied strongly with CNS age. Ancient CNSs were dominated by those conserved across all four genomes, followed by those retained only in progenitors (**Fig. 4D** and **Fig. S10**). In contrast, lineage-specific CNSs were most frequently conserved exclusively in progenitors, with the second most category being those conserved across all four genomes (**Fig. 4E** and **Fig. S10**). The next predominant groups within lineage-specific CNSs included CNSs restricted to progenitor A or progenitor B, as well as CNSs absent in one subgenome but present in the other three genomes (**Fig. 4E** and **Fig. S10**). These categories were also among the next common for ancient CNS (**Fig. 4D** and **Fig. S10**); however, progenitor-restricted CNSs were substantially more frequent than subgenome-missing CNSs among lineage-specific CNSs (**Fig. 4D** and **Fig. S10**), whereas this disparity was minimal for ancient CNSs (**Fig. 4E**

317 and **Fig. S10**). These patterns indicate that CNSs of different evolutionary ages follow distinct
318 trajectories. Ancient CNSs tend to be stably retained, whereas lineage-specific CNSs exhibit
319 substantial turnover. On the other hand, the large number of lineage-specific CNSs present in
320 one or both progenitors but absent from the polyploid (**Fig. 4E** and **Fig. S10**), suggests
321 widespread loss of these elements after hybridization and polyploidization. This may reflect the
322 inherent dynamism of *cis*-regulatory elements: some may rapidly evolve into new regulatory
323 sequences and diverge enough to escape detection as a lineage-specific CNSs. Alternatively,
324 hybridization, whole-genome duplication, and subsequent regulatory reorganization may drive
325 the extensive loss of lineage-specific CNSs. Further work will be needed to disentangle these
326 possibilities. We also observed many CNSs conserved exclusively between each subgenome and
327 its corresponding progenitors, particularly for lineage-specific CNSs (**Fig. 4D-4E** and **Fig. S10**),
328 further underscoring the strong influence of parental legacy on regulatory architecture. Other
329 categories occurred at much lower frequencies (**Fig. 4D-4E** and **Fig. S10**).
330
331 **Impacts of the CNS dynamics on chromatin accessibility variation of sequence-conserved ACRs**
332 To explore whether CNS turnover correlates with the chromatin accessibility variation observed
333 in sequence-conserved ACRs, we examined CNSs dynamics within these regions using a
334 comparative framework. We applied the same presence/absence–based CNS classification
335 described above (**Fig. 4D-4E** and **Fig. S10**), but in a more targeted context: each CNS was
336 categorized based on its association with genes in the same orthogroup across the four
337 genomes and its location within a sequence-conserved ACR and/or its corresponding BLASTed
338 regions (see Methods) (**Table S14**). We then summarized the distribution of CNS conservation
339 states for each ACR using a rank-normalized approach that accounts for both the CNS
340 conservation state and its abundance (see Methods).
341
342 The patterns were consistent between leaf and root tissues (**Fig. 5** and **Fig. S11**). ACRs with
343 parental subgenome-dominant chromatin accessibility were frequently associated with CNSs
344 conserved exclusively between that subgenome and its diploid progenitor (e.g., P_B-S_B in **Fig. 5**).
345 ACRs with parental no bias chromatin accessibility tended to contain CNSs conserved across all
346 four genomes (e.g., P_A-P_B-S_A-S_B in **Fig. 5**). ACRs exhibiting novel bias subgenome-dominant
347 chromatin accessibility were more likely to harbor CNSs unique to that subgenome (e.g., S_B in
348 **Fig. 5**) or CNSs missing in the other genomes (e.g., P_A-P_B-S_B in **Fig. 5**). In contrast, CNSs conserved
349 across three genomes were often not enriched in any chromatin accessibility category (e.g., P_B-
350 S_A-S_B, P_A-S_A-S_B, P_A-P_B-S_A in **Fig. 5**). CNSs conserved exclusively in progenitor A (e.g., P_A in **Fig. 5**) or
351 progenitor B (e.g., P_B in **Fig. 5**) were unexpectedly common across nearly all accessibility groups,
352 despite they being lost in the subgenomes. CNS exclusively conserved in both progenitors (e.g.,
353 P_A-P_B in **Fig. 5**) showed a distinctive pattern: they were broadly enriched but specifically
354 depleted in parental subgenome-dominant chromatin accessibility, suggesting that these ACRs
355 carry regulatory features that differ markedly between the two progenitor lineages. Together,
356 these results suggest that CNSs with different conservation states contribute meaningfully to
357 chromatin accessibility patterns of sequence-conserved ACRs in polyploid genome, shaping their
358 regulatory activities.
359
360 We also analyzed lineage- and polyploid-specific ACRs using simplified CNS classifications. For
361 lineage-specific ACRs, CNSs were categorized as conserved exclusively in the subgenome, in the
362 corresponding progenitor, or in both. The distributions of the CNS with different conservation

363 categories did not differ between lineage-specific ACRs with the same versus different
364 chromatin accessibility states (**Table S15** and **Fig. S12**). For polyploid-specific ACRs, CNSs were
365 categorized as conserved exclusively in one subgenome or in both. Among polyploid-specific
366 ACRs with differential accessibility, CNSs conserved in both subgenomes were consistently
367 depleted (**Table S16** and **Fig. S13**).
368
369 **Homoeolog expression bias in peanut and its underlying potential mechanisms**
370 To characterize homoeolog expression bias in peanut, we performed RNA-seq on leaf and root
371 tissues from the same three peanut species, each with two biological replicates. To enable cross-
372 genome comparisons, we focused on homoeologous genes identified across the two
373 subgenomes and their corresponding diploid progenitors. PCA of normalized gene counts
374 revealed that PC1 (44.3% of total variance) separated the diploid progenitors from the
375 subgenomes, PC2 (26%) further distinguished the two subgenomes, and PC3 (19.4%)
376 differentiated the two tissue types (**Fig. 6A**). Homoeologous genes displayed pronounced
377 expression divergence both between subgenomes and relative to their diploid counterparts,
378 whereas the two diploid progenitors showed highly similar expression profiles for these genes
379 (**Fig. 6A**).
380
381 Using the same conceptual framework applied to classify chromatin accessibility patterns we
382 categorized homoeolog expression bias (**Fig. 3B**). The overall distributions of these categories
383 were highly consistent between leaf and root tissues **(Table S17** and **Fig. 6B**). Most genes
384 exhibited no expression bias between the two subgenomes, with the majority showing a
385 parental no bias pattern (**Table S17** and **Fig. 6B**), consistent with previous findings (44). Among
386 genes exhibiting bias between the two subgenomes, novel bias accounted for the largest
387 proportion (~20%), followed by parental-legacy effects (~6%), and novel bias reverse (~3%)
388 (**Table S17** and **Fig. 6B**). These patterns suggest that most expression asymmetries might have
389 arisen *de novo* following hybridization and polyploidization, rather than being inherited from the
390 ancestral diploid progenitors.
391
392 To investigate how ACRs with different evolutionary trajectories relate to homoeolog expression
393 bias, we performed enrichment analyses linking expression categories with ACR evolutionary
394 types. Homoeolog expression patterns were simplified into four classes: no-bias, novel-bias,
395 parental-legacy bias and others. ACRs were first grouped by sequence conservation as peanut-
396 conserved, potential peanut-conserved, lineage-specific, polyploid-specific and subgenome-
397 specific and peanut-conserved ACRs were further subdivided by chromatin accessibility bias into
398 no-bias, novel-bias, parental-legacy bias and others. To ensure robust evolutionary inference,
399 we restricted the analysis to homoeologous genes associated with ACRs that could be
400 confidently classified based on BLAST conservation, excluding those had not been BLASTed due
401 to their absence from conserved syntenic regions. This integrative framework allowed us to
402 systematically examine whether specific classes of regulatory elements are preferentially
403 associated with distinct modes of homoeolog expression divergence.
404
405 To minimize the uncertainty in assigning ACRs to their target genes, we performed an additional
406 enrichment analysis excluding distal ACRs. Both the full and filtered datasets showed
407 comparable enrichment patterns, and the overall trends were consistent between leaf and root
408 tissues (**Table S18** and **Fig. S12**). Although few categories reached statistical significance, the

direction and magnitude of Z-scores offered meaningful trends (**Table S18** and **Fig. S12**). Given current power and multiple testing restrictions, we interpret these associations as trend-level and hypothesis-generating, pending expanded sampling and covariate-aware models. Genes showing no-bias expression were positively associated with peanut-conserved ACRs exhibiting no-bias in chromatin accessibility, whereas biased expression categories tended to show negative enrichment, particularly parental-legacy bias expression (**Fig. 6C** and **Table S18**). In contrast, peanut-conserved ACRs with novel-bias chromatin accessibility were positively enriched for genes with novel-bias expression but negatively or weakly enriched for no-bias and parental-legacy bias expression (**Fig.6C** and **Table S18**). Subgenome-specific ACRs exhibited similar tendencies, displaying negative enrichment with no-bias expression but positive enrichment with novel-bias and parental-legacy bias expression (**Fig. 6C** and **Table S18**). Lineage-specific ACRs and peanut-conserved ACRs with parental-legacy bias were generally positively enriched across all expression categories, with lineage-specific ACRs showing the strongest and statistically significant association with parental-legacy bias expression (**Fig. 6C** and **Table S18**). Conversely, polyploid-specific ACRs displayed consistently negative enrichment across all expression classes, particularly for no-bias expression (**Fig. 6C** and **Table S18**).

Together, these results indicate that genes with stable, unbiased expression tend to associate with conserved regulatory elements maintaining no-bias chromatin accessibility, whereas genes with biased expression shows enrichment for more recently evolved or divergent ACRs, such as lineage-specific ACRs, subgenome-specific ACRs or peanut-conserved ACRs with novel-bias chromatin accessibility. Meanwhile, the correspondence between ACR evolutionary origin and expression biases, where parental-legacy bias expression associates with lineage-specific ACRs and novel bias expression aligns with subgenome-specific ACRs, reveals a coordinated regulatory evolution, highlighting how ACRs with different evolutionary histories contribute to expression asymmetry in polyploid genome.

Discussion

Genetic variation within *cis*-regulatory elements is a major source of phenotypic diversity and evolutionary innovation. Understanding its evolutionary dynamics is therefore essential for elucidating the genetic basis of natural and artificial selection, and for facilitating targeted manipulation of gene regulation (6, 10, 56, 57). Although regulatory sequences exhibit rapid turnover across distantly related plant lineages (12), deeply conserved noncoding elements have also persisted across species diverged over 300 million years (10). However, how *cis*-regulatory elements evolve among closely related species and on lineage-specific time scale remains poorly understood (10, 12–14). Here we address this gap by systematically examining the evolutionary trajectories of regulatory elements in allotetraploid peanut and its diploid progenitors (**Fig. 2A**).

Beyond associating genetic variation with expression divergence in polyploids (16–19, 22, 58), we further investigated the evolutionary origins of regulatory variation, offering an additional dimension to understand *cis*-regulatory dynamics underlying expression changes. We uncovered a strong correspondence between homoeolog expression bias and ACR variation at both sequence and accessibility levels, reflecting their distinct evolutionary histories. For example, parental-legacy bias expression was enriched for lineage-specific ACRs, whereas novel-bias expression tends to associate with subgenome-specific or peanut-conserved ACRs with novel-bias chromatin accessibility (**Fig. 6C** and **Table S18**). In contrast, peanut-conserved ACRs with no

455 bias chromatin accessibility were depleted for biased expression but enriched for unbiased
456 expression (**Fig. 6C** and **Table S18**). We also detected a pronounced parental-legacy effect on
457 regulatory evolution (17, 45, 59–61), supported by lineage-specific ACRs (**Fig. 2B**), peanut-
458 conserved ACRs exhibiting parental-legacy bias chromatin accessibility (**Fig. 3C**) and
459 conservation of CNSs between a subgenome and its corresponding progenitor. (**Fig. 4D-4E**).
460 However, how the strength of this parental-legacy effect shifts over time remains an important
461 open question. Collectively, these findings highlight that ACSs with distinct evolutionary
462 trajectories exert differential impacts on regulatory landscapes in the polyploid genome.
463
464 From a broader evolutionary perspective, we characterized the fates of regulatory regions. Most
465 ACRs were sequence-conserved (**Fig. 2B**), exhibiting exceptionally high sequence similarity
466 (mean BLASTed ratio ~0.98 and mean mismatch rate ~0.02) (**Fig. 2D**), despite the ~2 Mya
467 divergence between the diploid progenitors (28). Many of these sequence-conserved ACRs also
468 maintained comparable histone modification profiles (**Fig. 2E**) and chromatin accessibility (**Fig.
469 3C**), likely reflecting both the recent formation of the polyploid and strong selective constrains
470 on regulatory function (10, 49, 50). Meanwhile, we also observed the emergence of new ACRs
471 that originated either *de novo* or through accumulated mutations in preexisting regulatory
472 sequences (6, 9–11). Polyploid-specific or subgenome-specific ACRs exemplify *de novo* origins,
473 arising after hybridization or after polyploidization (**Fig. 2A**). In contrast, some lineage-specific
474 ACRs likely represent gradual sequence divergence: ACRs that shift from peanut-conserved to
475 lineage-specific when the BLAST similarity threshold increases from 0.5 to 0.8 (**Fig. 2B**) illustrate
476 this process: these regions likely existed prior to the divergence of the diploid progenitors and
477 subsequently accumulated mutations, increasing intra-lineage similarity to ~80% while reducing
478 inter-lineage similarity to ~50%. Peanut-conserved ACRs with high sequence similarity but
479 biased chromatin accessibility provide another example of diversification (**Fig. 5**). The generally
480 lower frequency of subgenome- or polyploid-specific ACRs, compared with the more abundant
481 lineage-specific ACRs or peanut-conserved ACRs with biased chromatin accessibility, supports
482 the conclusion that regulatory diversification through gradual mutation accumulation is more
483 common than *de novo* emergence (10, 11, 22).
484
485 Importantly, we found that even minor sequence changes can have disproportionately large
486 effects on regulatory activity. Peanut-conserved ACRs with biased chromatin accessibility
487 exemplify this, where they exhibited high sequence similarity but substantial chromatin
488 accessibility differences that often manifest as presence/absence patterns rather than gradual
489 quantitative changes (**Fig. 3D**). Furthermore, we demonstrate a correlation between changes in
490 CNS composition and chromatin accessibility bias, a relationship not previously examined in
491 lineage-specific contexts (10). For instance, the peanut-conserved ACRs with parental
492 subgenome-dominant bias chromatin accessibility always possessed more CNSs that conserved
493 between the respective subgenome and its corresponding diploid, whereas peanut-conserved
494 ACRs with parental no bias chromatin accessibility were enriched for CNSs that conserved across
495 four genomes (**Fig. 5**). Although these associations are compelling, they remain correlative.
496 Without allele-specific assays or targeted perturbations, we cannot ascribe causality to CNS
497 turnover, and trans-acting factors, peak-calling thresholds, or tissue composition may also
498 contribute (62–64). A limitation is inherent to CNS identification itself, which requires a
499 predefined degree of sequence conservation. The large number of CNSs conserved exclusively in
500 one or both progenitors but absent from the subgenomes (**Fig. 4D-4E**) may suggest that these

501 sequences have undergone functional divergence in the subgenomes, shifting toward new
502 regulatory roles at the expense of detectable sequence conservation, even if the underlying
503 sequence changes might be minor. On the other hand, examining transcription factor binding
504 motifs (65) or footprinting data (56, 66) will also be essential for understanding how conserved
505 elements transition into novel regulatory functions and how frequently such events occur.
506
507 In summary, our study provides empirical evidence illuminating the evolutionary dynamics of
508 regulatory elements in polyploid plants. By integrating sequence variation, CNSs turnover,
509 chromatin accessibility, histone modification, and expression divergence, we provide a
510 comprehensive view of how *cis*-regulatory evolution drives expression diversification and
511 contributes to evolutionary innovation in plant genomes.
512
513 **Materials and Methods**
514 **Plant materials**
515 Three Arachis species were used in this study: the cultivated allotetraploid peanut *A. hypogaea*
516 (AABB, 2n=4x=40) and its two diploid progenitors, *A. duranensis* (AA, 2n=2x=20) and *A.*
517 *ipaensis* (BB, 2n=2x=20). Plants were grown in soil for approximately 10 days at 25 °C under
518 photoperiodic lighting (16 h light–8 h dark). Mature leaf and root tissues were collected with
519 two biological replicates per tissue for each species.
520
521 **ATAC-seq library preparation**
522 Nuclei isolation and purification were performed as described previously with modifications
523 (67). Then, the nuclei quality and density were evaluated with a hemocytometer under a
524 microscope. The ATAC-seq library amplification was performed following the protocol described
525 previously (68). About 50k nuclei were incubated with 2 µl Tn5 transposomes in 40 µl
526 tagmentation buffer at 37 °C for 30 min without rotation. The integration products were purified
527 using a Qiagen MinElute PCR Purification Kit or NEB Monarch DNA Cleanup Kit and then
528 amplified using Phusion DNA polymerase for 10–13 cycles. Amplified libraries were purified
529 using AMPure beads to remove primers.
530
531 **RNA-seq library preparation**
532 The collected tissues were flash-frozen with liquid N2 immediately after collection and stored at
533 -80C frezeer. The samples were ground to a powder using a pestle and mortar in liquid N2. Total
534 RNA was extracted and purified using TRIzol Reagent (Thermo Fisher Scientific) following the
535 manufacturer's instructions. For each sample, 1.3 µg of total RNA was prepared for sequencing
536 using the Illumina TruSeq mRNA Stranded Library Kit (Illumina) following the manufacturer's
537 instructions.
538
539 **ChIP-seq library preparation**
540 ChIP was performed following the general protocol of (69). Approximately 0.1 g of tissue were
541 used for each experiment. The fresh tissue was chopped into 0.5 mm cross sections and
542 crosslinked in accordance with the referenced protocol. Samples were immediately flash-frozen
543 in liquid nitrogen after crosslinking. Nuclei were extracted and lysed in 300 µl of lysis buffer and
544 was sonicated on a Diagenode Bioruptor. Tubes were centrifuged at 12,000g for 5 min and
545 supernatants transferred to new tubes. Dynabeads Protein A (Thermo Fisher Scientific, cat. no.
546 10002D) were washed with ChIP dilution buffer and then rotated with antibodies (anti-

547 H3K4me3, Millipore-Sigma, 07–473; anti-H3K36me3, Abcam, ab9050; anti-H3K56ac, Millipore-
548 Sigma, 07–667). 100 µl of ChIP dilution buffer for at least 4 h at 4C. Sonicated chromatin was
549 diluted tenfold in ChIP dilution buffer. For all samples and replicates, 460 µl of diluted chromatin
550 was incubated with 750 µg of Dynabeads Protein A coated with 1.5 µg of antibody. Samples
551 were rotated at 4 °C overnight, then washed, reverse-crosslinked, and treated with proteinase K
552 in accordance with the referenced protocol. DNA was purified by a standard phenol–chloroform
553 extraction followed by ethanol precipitation. The final libraries were prepared following exactly
554 with the protocol (69).
555
556 **Sequencing information**
557 All libraries were sequenced using an Illumina NovaSeq 6000 S4 150PE. All the raw reads of
558 ATAC-seq, ChIP-seq and RNA were uploaded to the National Center for Biotechnology
559 Information (NCBI).
560
561 **ATAC-seq data processing and alignment**
562 Raw reads from 12 libraries were cleaned with Trimmomatic version 0.36 (70) using a sliding
563 window trimming method with default setting except the minimal length of the reads as 30 bp.
564 The resulting clean reads were mapped to the respective reference genomes, including *A.*
565 *hypogaea*, *A. duranensis* and *A. ipaensis* from Legume Information System
566 (www.legumeinfo.org/collections/arachis/) using BWA-MEM aligner (71). For *A. hypogaea*, the
567 clean reads were mapped to each subgenome separately and then retained those uniquely
568 mapped with mapping quality >30 for further analyses. In addition, reads mapped to a blacklist
569 of regions exhibiting Tn5 integration bias from Tn5-treated genomic DNA and potential
570 collapsed sequences in the reference (1-kb windows with greater than 4x coverage over the
571 genome-wide median using ChIP-seq input) were removed using BEDtools (72). The clonal
572 duplicated were removed using Picard (73). Only ACRs located in the chromosomes were
573 retained for the following analyses.
574
575 **ACR identification and characterization**
576 MACS2 (74) was used to identify ACRs with the following parameters: --extsize 150 --shift -75 --
577 slocal 150 --llocal 1500 --max-gap 75 --min-length 175 --qvalue 0.05. To refine the peak calling
578 and identify small peaks, we used the makewindows function in BEDtools to split the identified
579 ACRs. Finally, the refined peaks that overlapped or within 50 bp were merged using the
580 BEDtools merge function. Proportion of reads around transcription start site (TSS) was
581 calculated as the number of identified ACRs are overlapped with 1 kp upstream of TSS. Fraction
582 of reads in peak (FRiP) was calculated as the percentage of the sequenced reads that overlapped
583 with the peaks.
584
585 **ChIP-seq analysis**
586 Reads were cleaned and mapped to reference genome using the same pipeline as for ATAC-seq
587 analysis. Only uniquely mapped reads with a mapping quality greater than 30 were retained for
588 downstream analyses. To normalize sequencing depth and compare libraries to input data, we
589 used the bamCompare function from deepTools2 (75) with the following parameters: "--
590 normalizeUsing BPM --pseudocount 1 --binSize 10 –extendReads". Signal densities for the
591 histone modifications from 2kb upstream to 2kb downstream around the genes were calculated
592 using the "computeMatrix" function in "scale-regions" mode with the setting of "-bs 20 -b 2000

593 -a 2000 –regionBodyLength 5000". Peak calling was performed using MACS2 (74) with the H3
594 ChIP-seq samples as controls. For H3K4me3, and H3K56ac ChIP-seq samples, narrow peaks were
595 identified using the parameters -p 1e-3 -f BAM --keep-dup all. For H3K36me3 ChIP-seq samples,
596 broad peaks were identified using the parameters --broad --broad-cutoff 0.05. The overlap
597 between histone modification peaks and the ACRs or their corresponding BLASTed sequences
598 were performed using bedtoolds. The resulting Matrices were loaded to a custom R script to
599 generate metaplots and do cluster analyses following (12).
600
601 **Differential chromatin accessibility**
602 To examine quantitative differences in ACRs between the two tissue-types from the same
603 species, we generated four pseudo replicates by randomly sampling the equal number of clean
604 reads from the two biological replicates using a customized Python script. Normalized read
605 counts were then calculated for each peak, and DESeq2 (76) was used to identify regions with
606 significant differences in accessibility. Peaks with an adjusted P-value (FDR) < 0.05 and an
607 absolute $\log_2$ fold change > 1 were defined as differential ACRs.
608
609 The same analytical framework was used to examine accessibility bias across peanut-conserved
610 ACRs and their corresponding BLASTed regions in the reference genomes (**Fig. 2B**). Parental no
611 bias was defined as the absence of significant accessibility differences between the two diploids
612 and between the two subgenomes. Parental subgenome A or B dominant bias was assigned
613 when a subgenome and its corresponding progenitor both exhibited significantly higher
614 accessibility relative to the other pair. No bias in progeny referred to cases with no significant
615 difference between the two subgenomes, further divided into progenitor A- or B-dominant
616 when one progenitor showed significantly higher chromatin accessibility. Novel bias in progeny
617 was defined as subgenome A- or B-dominant accessibility in the absence of significant
618 differences between the progenitors. Finally, reverse bias in progeny (subgenome A- or
619 subgenome B- dominant) referred to cases where a subgenome showed significantly higher
620 chromatin accessibility, but its corresponding progenitor had significantly lower chromatin
621 accessibility.
622
623 **Sequence conservation of ACRs identified in polyploidy peanut genome**
624 To investigate the sequence ancestry of ACRs identified in the two subgenomes of polyploid
625 peanut, we adopted a BLAST-based approach following the methods described in (13). ACR
626 sequences were used as queries, and reference regions were defined as the genomic intervals
627 spanning the two orthologous genes flanking each ACR in the corresponding genomes (Fig. S2).
628 These reference regions were identified using GENESPACE (77) and classified based on
629 conservation of the orthologous gene pairs – either conserved across all four genomes (peanut-
630 conserved), conserved across three genomes or between a subgenome and its non-
631 corresponding diploid progenitor (potential peanut-conserved), exclusively conserved between
632 a subgenome and its corresponding diploid progenitor (lineage-specific), exclusively between
633 the two subgenomes (polyploid-specific), or conserved only in one subgenome (subgenome-
634 specific). ACRs from each subgenome were BLASTed against the reference regions using BLAST
635 (78) and the same parameters as described in (13). Sequence intersections were identified using
636 BEDtools (72) and sequence extractions were performed with seqKit (79). The resulting BLAST
637 outputs were then combined and filtered using customized scripts.
638

639 **Analyses of CNS dynamics in polyploid-diploid framework**
640 CNS information of three peanut species, *A. hypogaea*, *A. duranensis* and *A. ipaensis.* Was
641 downloaded from (https://conservatorycns.com/dist/pages/conservatory/index.php). We
642 summarized the counts, repeats, and conserved information for each CNSs across three peanut
643 species or four genomes using customed R script. When examining the CNSs located within the
644 sequence-conserved ACRs, we first used BEDtools to intersect the coordinates of these CNSs
645 with the coordinates of sequence-conserved ACRs or their corresponding BLASTed sequences,
646 with the requirement that the CNSs might be completely within the ACRs or BLASTed sequences
647 regions using the commend of -f 1.0. Then we analyzed all the CNSs within each ACR separately
648 and further grouped them by the genes they are associated with, requiring that genes in each
649 group belong to the same orthogroup. Subsequently, we categorized CNSs based on their
650 conservation states across four genomes – two subgenomes in polyploid peanut and two diploid
651 progenitors and divided them as: (i) conserved across all four genomes, (ii) conserved across
652 three genomes, (iii) conserved exclusively in progenitors, (iv) conserved exclusively in polyploid
653 subgenomes, (v) conserved between a subgenome and its corresponding progenitor, (vi)
654 conserved between a subgenome and the non-corresponding progenitor, and (vii) restricted to
655 a single genome.
656
657 To quantify the relative enrichment of conservation patterns across CNS categories within each
658 orthogroup within each ACR, we generated a rank-normalized matrix from the CNS
659 categorization count matrix. For each CNS, we first identified the categories with non-zero
660 counts. Categories with zero counts were treated as absent and assigned a rank-normalized
661 score of zero. For categories with positive values, we computed ranks based on descending
662 abundance, such that the highest count received rank 1 and the lowest received rank N, where
663 N is the number of present categories. These ranks were then normalized by dividing by N,
664 producing values between 1/N and 1. The resulting rank-normalized scores were placed back
665 into a full vector, with zeros retained for absent categories. This procedure was applied row-
666 wise to the CNS matrix, producing a final rank-normalized matrix with the same dimensions as
667 the original input.
668
669 **RNA-seq data processing and alignment**
670 The adapters and low-quality reads were trimmed using Timmomatic with the following
671 parameters: SLIDINGWINDOW: 4:15 LEADING:3 TRAILING:3 MINLEN:25. The clean reads were
672 mapped to reference geome using STAR (80). Gene expression is quantified using GeneCounts
673 mode and only uniquely mapped reads are counted (--outFilterMultimapNmax 1 --
674 outFilterMismatchNmax 999).
675
676 **Differential gene expression analysis**
677 Peptide sequences from the two subgenomes of *A. hypogaea* and the genomes of *A. duranensis*
678 and *A. ipaensis* were used to identify putative orthologs using OrthoFinder (81). Orthogroups
679 containing single-copy homologs in a 1:1:1:1 ratio across these four genomes were retained.
680 Variance-stabilizing transformed gene counts, normalized with DESeq2, were used for PCA. Leaf
681 and root tissues were analyzed separately, and only orthogroups with an average normalized
682 gene count greater than 5 in at least one genome were included in subsequent analyses.
683 Pairwise differential expression analyses were performed in 'DESeq2' package, with genes
684 showing an adjusted P-value (FDR) < 0.05 and an absolute $\log_2$ fold change > 1 considered

differentially expressed. The same criteria used to classify accessibility bias was applied to categorize homoeolog expression bias.

Enrichment between expression bias and ACR types

Only Orthogroups exclusively associated with ACRs that had been assigned an evolutionary origin were included in the analysis. Expression bias categories were simplified into four types: no bias (including "no bias in progeny" and "parental no bias"), novel bias (including "novel bias" and "novel bias reverse"), parental-legacy bias (including "parental subgenome A- dominant" and "parental subgenome B- dominant"), and others. Associated ACRs were first classified by sequence conservation into peanut-conserved, potential peanut-conserved, lineage-specific, polyploid-specific, and subgenome-specific. Peanut-conserved ACRs were further subdivided based on their simplified chromatin accessibility bias (analogous to expression bias) into no bias, novel bias, parental-legacy bias, and others. Permutation-based enrichment analyses were performed separately for leaf and root tissues using both unfiltered and filtered datasets (with distal ACRs removed).

For each Orthogroup, the proportion of ACRs belonging to each category was calculated by dividing the number of ACRs of that type by the total number of ACRs within that Orthogroup. These normalized proportions were then aggregated by expression bias groups. Observed mean proportions for each ACR category were calculated across Orthogroups within each expression group. To evaluate the significance of these observed proportions, 10,000 random permutations were conducted by shuffling expression bias group labels while preserving the ACR composition within each Orthogroup. For each permutation, mean proportions were recalculated to generate a null distribution. The null mean and standard deviation were then used to compute standardized Z-scores, and empirical P-values were derived as the proportion of permuted means equal to or greater than the observed mean proportion. Multiple testing correction was applied using the Benjamini–Hochberg (BH) method, and categories with adjusted $P < 0.05$ were considered significantly enriched.

Acknowledgments

We sincerely appreciate the members of the Schmitz lab for their valuable analytical suggestions. J.P.M. was supported by an NIH Training Grant (T32GM142623). This research was supported by the National Science Foundation (IOS-1856627) awarded to R.J.S.

References

1. A. P. Marand, A. L. Eveland, K. Kaufmann, N. M. Springer, cis-Regulatory Elements in Plant Development, Adaptation, and Evolution. *Annu. Rev. Plant Biol.* **74**, 111–137 (2023).

2. R. J. Schmitz, E. Grotewold, M. Stam, Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. *Plant Cell* **34**, 718–741 (2022).

3. S. L. Klemm, Z. Shipony, W. J. Greenleaf, Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).

4. P. J. Wittkopp, G. Kalay, Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics* **13**, 59–69 (2012).

728 5. H. K. Long, S. L. Prescott, J. Wysocka, Ever-Changing Landscapes: Transcriptional Enhancers
729 in Development and Evolution. *Cell* **167**, 1170–1187 (2016).

730 6. X. Li, R. J. Schmitz, Cis-regulatory dynamics in plant domestication. *Trends Genet.* **0** (2025).

731 7. P. S. Soltis, D. E. Soltis, Plant genomes: Markers of evolutionary history and drivers of
732 evolutionary change. *Plants People Planet* **3**, 74–82 (2021).

733 8. F. Murat, Y. Van de Peer, J. Salse, Decoding plant and animal genome plasticity from
734 differential paleo-evolutionary patterns and processes. *Genome Biol. Evol.* **4**, 917–928
735 (2012).

736 9. A. Katikaneni, C. B. Lowe, Novelty versus innovation of gene regulatory elements in human
737 evolution and disease. *Curr. Opin. Genet. Dev.* **90**, 102279 (2025).

738 10. K. R. Amundson, *et al.*, A deep-time landscape of plant *cis*-regulatory sequence evolution.
739 *bioRxiv* 2025.09.17.676453 (2025).

740 11. J. M. C. McDonald, R. D. Reed, Beyond modular enhancers: new questions in cis-regulatory
741 evolution. *Trends Ecol. Evol.* **0** (2024).

742 12. Z. Lu, *et al.*, The prevalence, evolution and chromatin signatures of plant regulatory
743 elements. *Nat Plants* **5**, 1250–1259 (2019).

744 13. J. P. Mendieta, *et al.*, Investigating the cis-regulatory basis of C3 and C4 photosynthesis in
745 grasses at single-cell resolution. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2402781121 (2024).

746 14. H. Yan, *et al.*, A single-cell rice atlas integrates multi-species data to reveal cis-regulatory
747 evolution. *Nat. Plants* **11**, 2050–2071 (2025).

748 15. E. Kuzmin, J. S. Taylor, C. Boone, Retention of duplicated genes in evolution. *Trends Genet.*
749 **38**, 59–72 (2022).

750 16. C. Fang, *et al.*, Dynamics of *cis*-regulatory sequences and transcriptional divergence of
751 duplicated genes in soybean. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2303836120 (2023).

752 17. G. Hu, *et al.*, Evolutionary Dynamics of Chromatin Structure and Duplicate Gene Expression
753 in Diploid and Allopolyploid Cotton. *Mol. Biol. Evol.* **41** (2024).

754 18. J. Han, *et al.*, Genome-wide chromatin accessibility analysis unveils open chromatin
755 convergent evolution during polyploidization in cotton. *Proc. Natl. Acad. Sci. U. S. A.* **119**,
756 e2209743119 (2022).

757 19. C. Fang, *et al.*, Dynamics of accessible chromatin regions and subgenome dominance in
758 octoploid strawberry. *Nat. Commun.* **15**, 1–14 (2024).

759 20. F. Almeida-Silva, Y. Van de Peer, Gene expression divergence following gene and genome
760 duplications in spatially resolved plant transcriptomes. *Plant Cell* **37**, koaf243 (2025).

761 21. T. Srikant, A. Gonzalo, K. Bomblies, Chromatin accessibility and gene expression vary
762 between a new and evolved autopolyploid of Arabidopsis arenosa. *Mol. Biol. Evol.* **41**,
763 msae213 (2024).

764 22. X. Li, X. Zhang, R. J. Schmitz, From duplication to divergence: single-cell insights into
765 transcriptional and cis-regulatory landscapes in soybean. *Plant Cell* koaf279 (2025).

766 23. G. Seijo, *et al.*, Genomic relationships between the cultivated peanut (Arachis hypogaea,
767 Leguminosae) and its close relatives revealed by double GISH. *Am. J. Bot.* **94**, 1963–1971
768 (2007).

769 24. G. Robledo, G. I. Lavia, G. Seijo, Species relations among wild Arachis species with the A
770 genome as revealed by FISH mapping of rDNA loci and heterochromatin detection. *Theor.*
771 *Appl. Genet.* **118**, 1295–1307 (2009).

772 25. D. J. Bertioli, *et al.*, The genome sequences of Arachis duranensis and Arachis ipaensis, the
773 diploid ancestors of cultivated peanut. *Nat. Genet.* **48**, 438–446 (2016).

774 26. X. Chen, *et al.*, Draft genome of the peanut A-genome progenitor (*Arachis duranensis*)
775 provides insights into geocarpy, oil biosynthesis, and allergens. *Proc. Natl. Acad. Sci. U. S. A.*
776 **113**, 6785–6790 (2016).

777 27. Q. Lu, *et al.*, Genome Sequencing and Analysis of the Peanut B-Genome Progenitor (Arachis
778 ipaensis). *Front. Plant Sci.* **9**, 604 (2018).

779 28. X. Chen, *et al.*, Sequencing of Cultivated Peanut, Arachis hypogaea, Yields Insights into
780 Genome Evolution and Oil Improvement. *Mol. Plant* **12**, 920–934 (2019).

781 29. K. Zhao, *et al.*, Pangenome analysis reveals structural variation associated with seed size
782 and weight traits in peanut. *Nat. Genet.* 1–12 (2025).

783 30. X. Zhang, *et al.*, A spatially resolved multi-omic single-cell atlas of soybean development.
784 *Cell* **188**, 550-567.e19 (2024).

785 31. A. Farmer, S. Thibivilliers, K. H. Ryu, J. Schiefelbein, M. Libault, Single-nucleus RNA and ATAC
786 sequencing reveals the impact of chromatin accessibility on gene expression in Arabidopsis
787 roots at the single-cell level. *Mol. Plant* **14**, 372–383 (2021).

788 32. M. W. Dorrity, *et al.*, The regulatory landscape of Arabidopsis thaliana roots at single-cell
789 resolution. *Nat. Commun.* **12**, 3334 (2021).

790 33. V. Sundaram, J. Wysocka, Transposable elements as a potent source of diverse cis-
791 regulatory sequences in mammalian genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**,
792 20190347 (2020).

793 34. A. Y. Du, J. D. Chobirko, X. Zhuo, C. Feschotte, T. Wang, Regulatory transposable elements in
794 the encyclopedia of DNA elements. *Nat. Commun.* **15**, 7594 (2024).

795 35. W. A. Ricci, *et al.*, Widespread long-range cis-regulatory elements in the maize genome. *Nat*
796 *Plants* (2019).

797 36. A. E. Yocca, Z. Lu, R. J. Schmitz, M. Freeling, P. P. Edger, Evolution of Conserved Noncoding
798 Sequences in Arabidopsis thaliana. *Mol. Biol. Evol.* **38**, 2692–2703 (2021).

799 37. L. Quadrana, I. R. Henderson, The natural history of transposons in plant pangenomes and
800 panepigenomes. *Curr. Opin. Plant Biol.* **88**, 102818 (2025).

801 38. S. Moser Tralamazza, L. Nanchira Abraham, C. S. Reyes-Avila, B. Corrêa, D. Croll, Histone
802 H3K27 methylation perturbs transcriptional robustness and underpins dispensability of
803 highly conserved genes in fungi. *Mol. Biol. Evol.* **39**, msab323 (2022).

804 39. G. Vernaz, *et al.*, Epigenetic divergence during early stages of speciation in an African crater
805 lake cichlid fish. *Nat. Ecol. Evol.* **6**, 1940–1951 (2022).

806 40. J. P. Mendieta, A. P. Marand, W. A. Ricci, X. Zhang, R. J. Schmitz, Leveraging histone
807 modifications to improve genome annotations. *G3* **11** (2021).

808 41. H. Le, C. H. Simmons, X. Zhong, Functions and mechanisms of histone modifications in
809 plants. *Annu. Rev. Plant Biol.* **76**, 551–578 (2025).

810 42. X. Zhang, Y. V. Bernatavichute, S. Cokus, M. Pellegrini, S. E. Jacobsen, Genome-wide analysis
811 of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. *Genome Biol.*
812 **10**, R62 (2009).

813 43. F. Roudier, *et al.*, Integrative epigenomic mapping defines four main chromatin states in
814 Arabidopsis: Organization of the Arabidopsis epigenome. *EMBO J.* **30**, 1928–1938 (2011).

815 44. H. Yu, *et al.*, Epigenetic insights into the domestication of tetraploid peanut. *Plant Physiol.*
816 **198**, kiaf254 (2025).

817 45. M.-J. Yoo, E. Szadkowski, J. F. Wendel, Homoeolog expression bias and expression level
818 dominance in allopolyploid cotton. *Heredity (Edinb.)* **110**, 171–180 (2013).

819 46. J. Wu, *et al.*, Homoeolog expression bias and expression level dominance in resynthesized
820 allopolyploid Brassica napus. *BMC Genomics* **19**, 586 (2018).

821 47. C. E. Grover, *et al.*, Homoeolog expression bias and expression level dominance in
822 allopolyploids. *New Phytol.* **196**, 966–971 (2012).

823 48. M. Li, R. Wang, X. Wu, J. Wang, Homoeolog expression bias and expression level dominance
824 (ELD) in four tissues of natural allotetraploid Brassica napus. *BMC Genomics* **21**, 330 (2020).

825 49. A. Hendelman, *et al.*, Conserved pleiotropy of an ancient plant homeobox gene uncovered
826 by cis-regulatory dissection. *Cell* (2021).

827 50. A. Lanctot, A. Hendelman, P. Udilovich, G. M. Robitaille, Z. B. Lippman, Antagonizing cis-
828 regulatory elements of a conserved flowering gene mediate developmental robustness.
829 *Proc. Natl. Acad. Sci. U. S. A.* **122**, e2421990122 (2025).

830 51. D. Ciren, S. Zebell, Z. B. Lippman, Extreme restructuring of cis-regulatory regions controlling
831 a deeply conserved plant stem cell regulator. *PLoS Genet.* **20**, e1011174 (2024).

832 52. X. Wang, *et al.*, Dissecting cis-regulatory control of quantitative trait variation in a plant
833 stem cell circuit. *Nat Plants* **7**, 419–427 (2021).

834 53. T. C. Tran, *et al.*, Altered interactions between cis-regulatory elements partially resolve
835 BLADE-ON-PETIOLE genetic redundancy in Capsella rubella. *Plant Cell* **36**, 4637–4657 (2024).

836 54. Y. Luo, *et al.*, Characterization and functional analysis of conserved non-coding sequences
837 among poaceae: insights into gene regulation and phenotypic variation in maize. *BMC*
838 *Genomics* **26**, 46 (2025).

839 55. J. Van de Velde, M. Van Bel, D. Vaneechoutte, K. Vandepoele, A collection of conserved
840 noncoding sequences to study gene regulation in flowering plants. *Plant Physiol.* **171**, 2586–
841 2598 (2016).

842 56. A. P. Marand, *et al.*, The genetic architecture of cell type-specific cis regulation in maize.
843 *Science* **388**, eads6601 (2025).

844 57. D. Villar, P. Flicek, D. T. Odom, Evolution of transcription factor binding in metazoans -
845 mechanisms and functional implications. *Nat. Rev. Genet.* **15**, 221–233 (2014).

846 58. M. Li, W. Sun, F. Wang, X. Wu, J. Wang, Asymmetric epigenetic modification and homoeolog
847 expression bias in the establishment and evolution of allopolyploid Brassica napus. *New*
848 *Phytol.* **232**, 898–913 (2021).

849 59. K. A. Bird, *et al.*, Replaying the evolutionary tape to investigate subgenome dominance in
850 allopolyploid Brassica napus. *New Phytol.* (2020).

851 60. P. P. Edger, *et al.*, Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* **51**,
852 541–547 (2019).

853 61. R. J. A. Buggs, *et al.*, The legacy of diploid progenitors in allopolyploid gene expression
854 patterns. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130354 (2014).

855 62. M. A. A. Minow, *et al.*, Hybrids reveal accessible chromatin trans genetic associations.
856 *bioRxivorg* 2025.10.09.681407 (2025).

857 63. M. A. Ballinger, K. L. Mack, S. M. Durkin, E. A. Riddell, M. W. Nachman, Environmentally
858 robust cis-regulatory changes underlie rapid climatic adaptation. *Proc. Natl. Acad. Sci. U. S.*
859 *A.* **120**, e2214614120 (2023).

860 64. M. Lensink, G. Monroe, D. J. Kliebenstein, Trans-regulatory loci shape natural variation of
861 gene expression plasticity in Arabidopsis. *Genetics* **230**, iyaf116 (2025).

862 65. L. A. Baumgart, *et al.*, Recruitment, rewiring and deep conservation in flowering plant gene
863 regulation. *Nat. Plants* **11**, 1514–1527 (2025).

864 66. Y. Hu, *et al.*, Multiscale footprints reveal the organization of cis-regulatory elements. *Nature*
865 **638**, 779–786 (2025).

866 67. X. Zhang, A. P. Marand, H. Yan, R. J. Schmitz, scifi-ATAC-seq: massive-scale single-cell
867 chromatin accessibility sequencing using combinatorial fluidic indexing. *Genome Biol.* **25**, 90
868 (2024).

869 68. Z. Lu, B. T. Hofmeister, C. Vollmers, R. M. DuBois, R. J. Schmitz, Combining ATAC-seq with
870 nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res.* **45**,
871 e41 (2017).

872 69. W. A. Ricci, L. Levin, X. Zhang, Genome-wide profiling of histone modifications with ChIP-
873 seq. *Methods Mol. Biol.* **2072**, 101–117 (2020).

874 70. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence
875 data. *Bioinformatics* **30**, 2114–2120 (2014).

876 71. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*
877 *[q-bio.GN]* (2013).

878 72. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic
879 features. *Bioinformatics* **26**, 841–842 (2010).

880 73. Picard toolkit. *Broad Institute, GitHub repository* [Preprint] (2019). Available at:
881 https://broadinstitute.github.io/picard/.

882 74. Y. Zhang, *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

883 75. F. Ramírez, *et al.*, deepTools2: a next generation web server for deep-sequencing data
884 analysis. *Nucleic Acids Res.* **44**, W160-5 (2016).

885 76. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for
886 RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

887 77. J. T. Lovell, *et al.*, GENESPACE tracks regions of interest and gene copy number variation
888 across multiple genomes. *Elife* **11** (2022).

889 78. C. Camacho, *et al.*, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421
890 (2009).

891 79. W. Shen, S. Le, Y. Li, F. Hu, SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File
892 Manipulation. *PLoS One* **11**, e0163962 (2016).

893 80. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

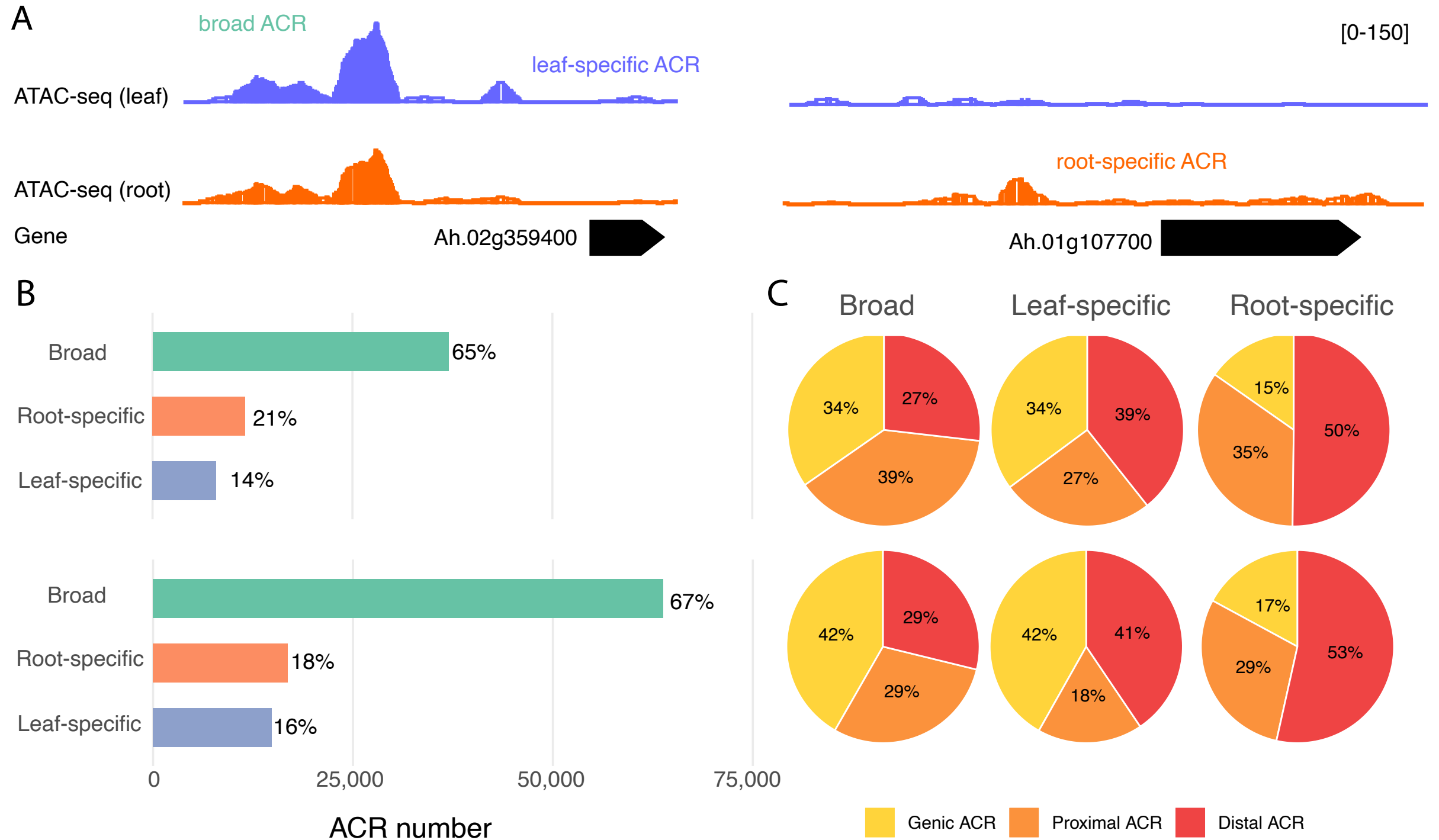
894 81. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative
895 genomics. *Genome Biol.* **20**, 238 (2019).

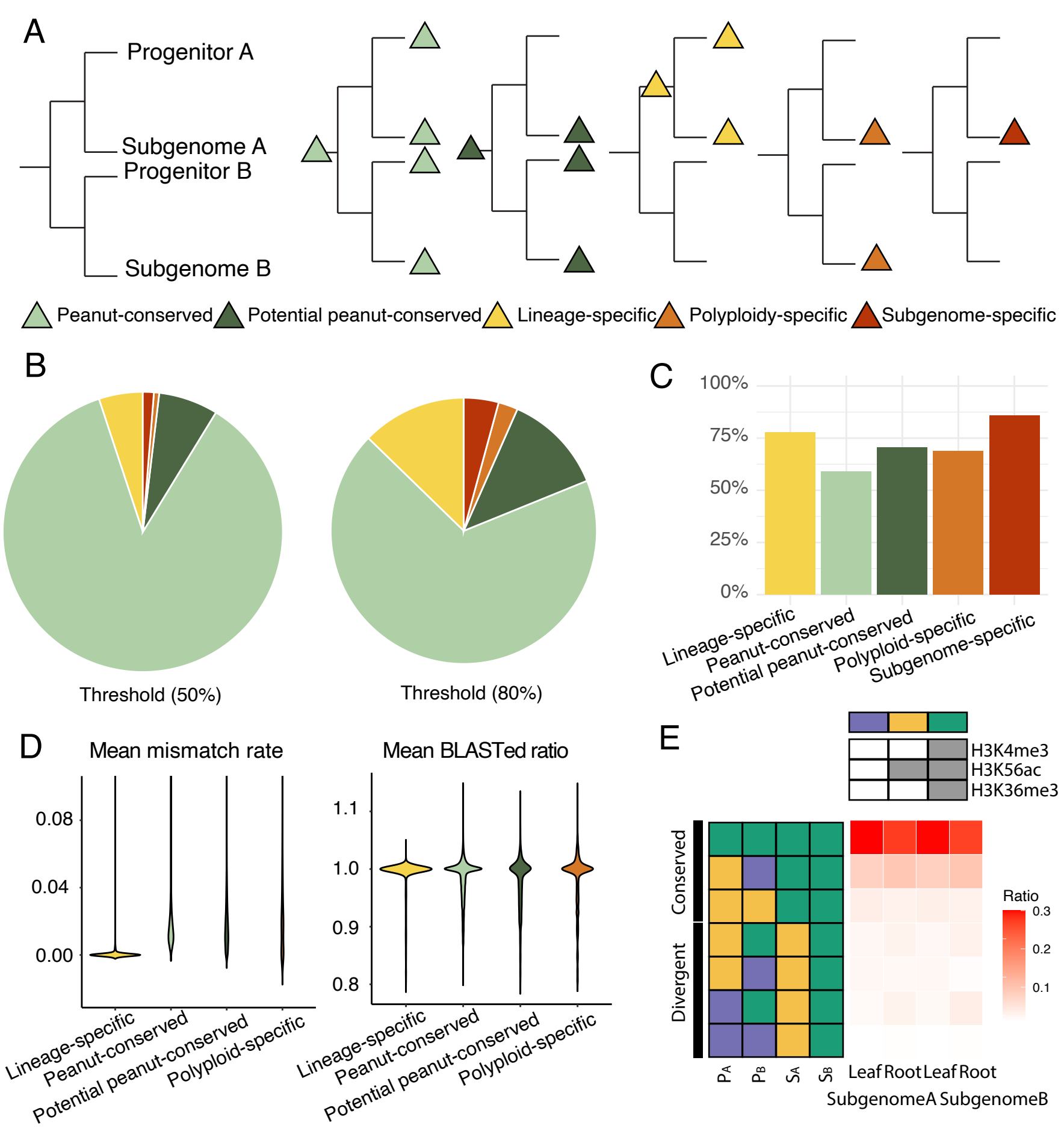
896

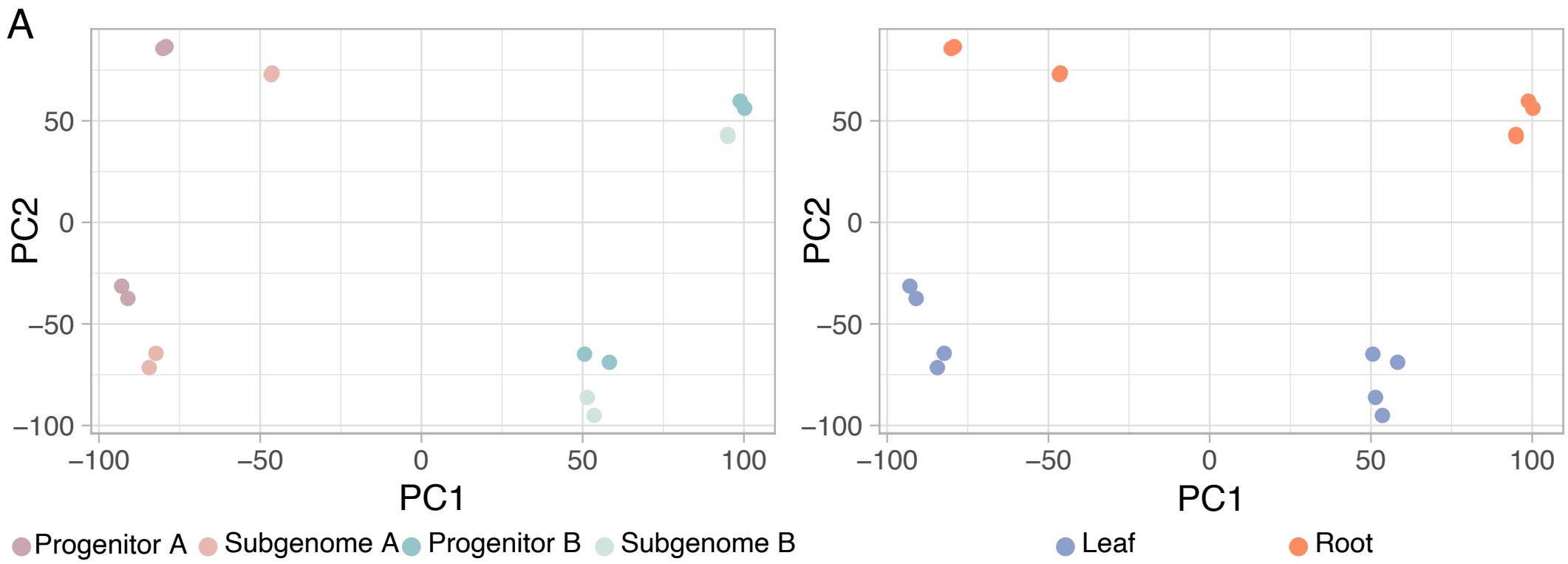
897 **Figure captions**
898 **Figure 1.** Characterization of ACRs identified in peanut and diploid progenitor.
899 **(A)** Screenshot illustrating example of tissue-specific ACRs and broad ACRs. **(B)** Number and
900 proportion of tissue-specific versus broad ACRs in subgenome B of *A. hypogaea* (top) and
901 progenitor B (*A. ipaensis*) (bottom). **(C)** Distribution of genic, proximal and distal ACRs within
902 broad, leaf-specific and root-specific categories in subgenome B (top) and progenitor genome B
903 (bottom).
904
905 **Figure 2.** Identification and characterization of ACRs with distinct evolutionary trajectories.
906 **(A)** Schematic illustrating the classification of ACRs based on sequence conservation profiles.
907 Peanut-conserved ACRs are retained across all four genomes, the two subgenomes of polyploid
908 peanut and both diploid progenitors, reflecting ancestral conservation predating progenitor
909 divergence. Potential peanut-conserved ACRs are detected in three genomes, or between a
910 subgenome and a non-corresponding progenitor. Lineage-specific ACRs are shared between a
911 subgenome and its corresponding progenitor, representing inheritance from the respective
912 diploid lineage. Polyploid-specific ACRs likely originated after hybridization but before whole-
913 genome duplication. Subgenome-specific ACRs are unique to one subgenome, suggesting *de*
914 *novo* emergence following polyploidization. **(B)** Distribution of ACRs with distinct evolutionary
915 trajectories (identified in subgenome A) under two sequence-similarity thresholds, using
916 BLASTed ratios ≥50% (left) and ≥80% (right) to define sequence conservation. **(C)** Proportion of
917 each evolutionary ACR category overlapping with TEs (in subgenome B), with the criteria of at
918 least 20 bp overlap. **(D)** Genetic variation across peanut-conserved, potential peanut-conserved,
919 lineage-specific, and polyploid-specific ACRs (in subgenome B), measured by mean mismatch
920 rate and mean BLASTed ratio. **(E)** Distribution of distinct histone modification patterns across
921 conserved ACRs and their corresponding BLASTed regions in four genomes, the two
922 subgenomes of polyploid peanut and both diploid progenitors. The top panel summarizes the
923 three histone modification states (purple, blue and green bars) defined by combinations of the
924 histone modification marks, H3K4me3, H3K56ac and H3K36me3, with grey bars indicating the
925 presence of each respective mark. The bottom-left schematic illustrates the classification of
926 histone modification patterns shown from top to bottom: (i) patterns conserved across all four
927 genomes; (ii) patterns shared between the two subgenomes but different in both progenitors;
928 (iii) patterns shared between both subgenomes and both progenitors but differing between
929 subgenome and progenitor groups; (iv) patterns distinct between the two subgenomes but each
930 matching its corresponding progenitor; (v) patterns consistent only between subgenome A and
931 progenitor A; (vi) patterns consistent only between subgenome B and progenitor B; and (vii)
932 patterns different from both progenitors. The top three categories represent conserved histone
933 modification patterns between the two subgenomes, whereas the remaining four represent
934 divergent patterns. The bottom-right heatmap shows the proportion of each histone
935 modification pattern in leaf and root tissues within subgenome A and subgenome B. P_A:
936 progenitor A; P_B: progenitor B; S_A: subgenome A and S_B: subgenome B.
937
938 **Figure 3**. Characterization of chromatin accessibility bias among peanut-conserved ACRs.
939 **(A)** Principal component analyses (PCA) of normalized chromatin accessibility counts for peanut-
940 conserved ACRs (in subgenome A) across four genome types: two subgenomes and their
941 corresponding diploid progenitors (left) and across two tissue types: leaf and root (right). **(B)**
942 Schematic illustrating classification of chromatin accessibility bias based on the relative

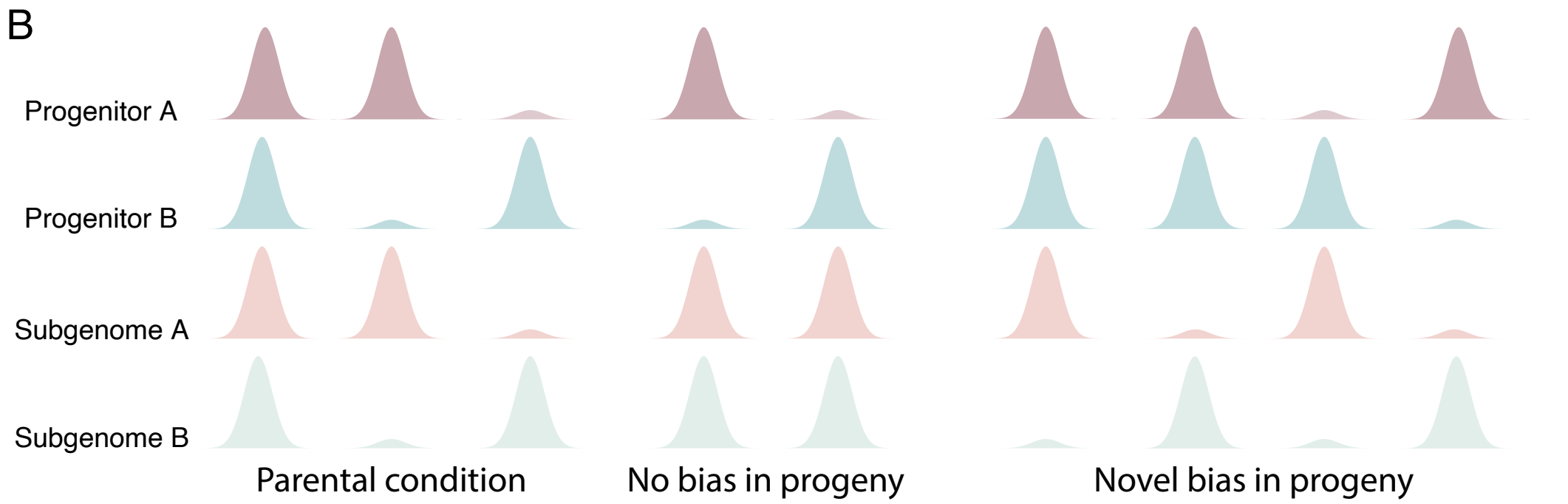
943 chromatin accessibility patterns between subgenomes and between their corresponding
944 progenitors. Categories include (from left to right): parental no bias (no bias between either
945 subgenomes or progenitors); parental subgenome A- dominant or subgenome B- dominant (bias
946 direction in subgenome consistent with that in progenitors); no bias progenitor A- dominant or
947 progenitor B-dominant (no bias between subgenomes but bias present between progenitors);
948 novel bias subgenome A- or subgenome B- dominant (bias emerging between subgenomes
949 despite no bias between progenitors); and novel bias-reverse (novel bias -r) subgenome A- or
950 subgenome B-dominant (bias direction opposite the observed between progenitors). **(C)**
951 Distribution of these chromatin accessibility bias categories for peanut-conserved ACRs
952 identified in subgenome A and subgenome B across leaf and root tissues. **(D)** Overlap rate
953 between BLASTed regions of peanut-conserved ACRs (identified in subgenome B) in subgenome
954 A and the ACRs identified in subgenome A; red labels denote subgenome B-dominant ACRs. P_A:
955 progenitor A; P_B: progenitor B; S_A: subgenome A and S_B: subgenome B.
956
957 **Figure 4.** Characterization of CNS dynamics in the polyploid-diploid peanut system.
958 **(A)** Number and proportion of CNSs identified in the three peanut species, grouped by their
959 evolutionary age of conservation, from ancient Tracheophyte-level or Angiosperm-level CNSs to
960 increasingly lineage-restricted CNSs in Dicots, Rosids, and Fabaceae. **(B)** Distribution of CNS copy
961 numbers across CNSs with different evolutionary age of conservation. **(C)** Distribution of CNSs
962 with different conservation levels across the three peanut species (left) and across the four
963 genomes, the two subgenomes of polyploid peanut and the two diploid progenitors (right), for
964 each evolutionary age of conservation category. Numbers from 1 to 4 indicate CNSs conserved
965 in one, two, three, or all four species/genomes. **(D-E)** Complex spectrum of CNS dynamics (CNS
966 conservation categories) across the diploid-polyploid framework for Angiosperm-level CNSs (D)
967 and Fabaceae-level CNSs (E). CNSs are categorized as: (i) conserved across all four genomes, (ii)
968 conserved across three genomes, (iii) conserved exclusively in progenitors, (iv) conserved
969 exclusively in polyploid subgenomes, (v) conserved between a subgenome and its corresponding
970 progenitor, (vi) conserved between a subgenome and the non-corresponding progenitor, and
971 (vii) restricted to a single genome. P_A: progenitor A; P_B: progenitor B; S_A: subgenome A and S_B:
972 subgenome B.
973
974 **Figure 5.** Rank-normalized scores of CNS conservation categories across ACRs with different
975 chromatin accessibility bias.
976 Heatmap showing the rank-normalized scores of CNS conservation patterns (as defined in Fig.
977 4D-4E) across peanut-conserved ACRs in subgenome B, annotated according to leaf chromatin
978 accessibility categories (as in Fig. 3C). For each ACR, CNS conservation categories were
979 summarized using a rank-normalization approach: all non-zero CNS conservation categories
980 were ranked in descending order of abundance and normalized by the number of categories
981 present, producing values between 0 and 1. A value of 1 indicates that an ACR contains CNSs
982 exclusively from a single conservation category. Rows are grouped by ACR chromatin
983 accessibility categories (as in Fig. 3C), and the color scale reflects the relative rank-normalized
984 score within each ACR. P_A: progenitor A; P_B: progenitor B; S_A: subgenome A and S_B: subgenome
985 B.
986
987 **Figure 6.** Homeolog expression bias in peanut and its association with ACRs of distinct
988 evolutionary trajectories.

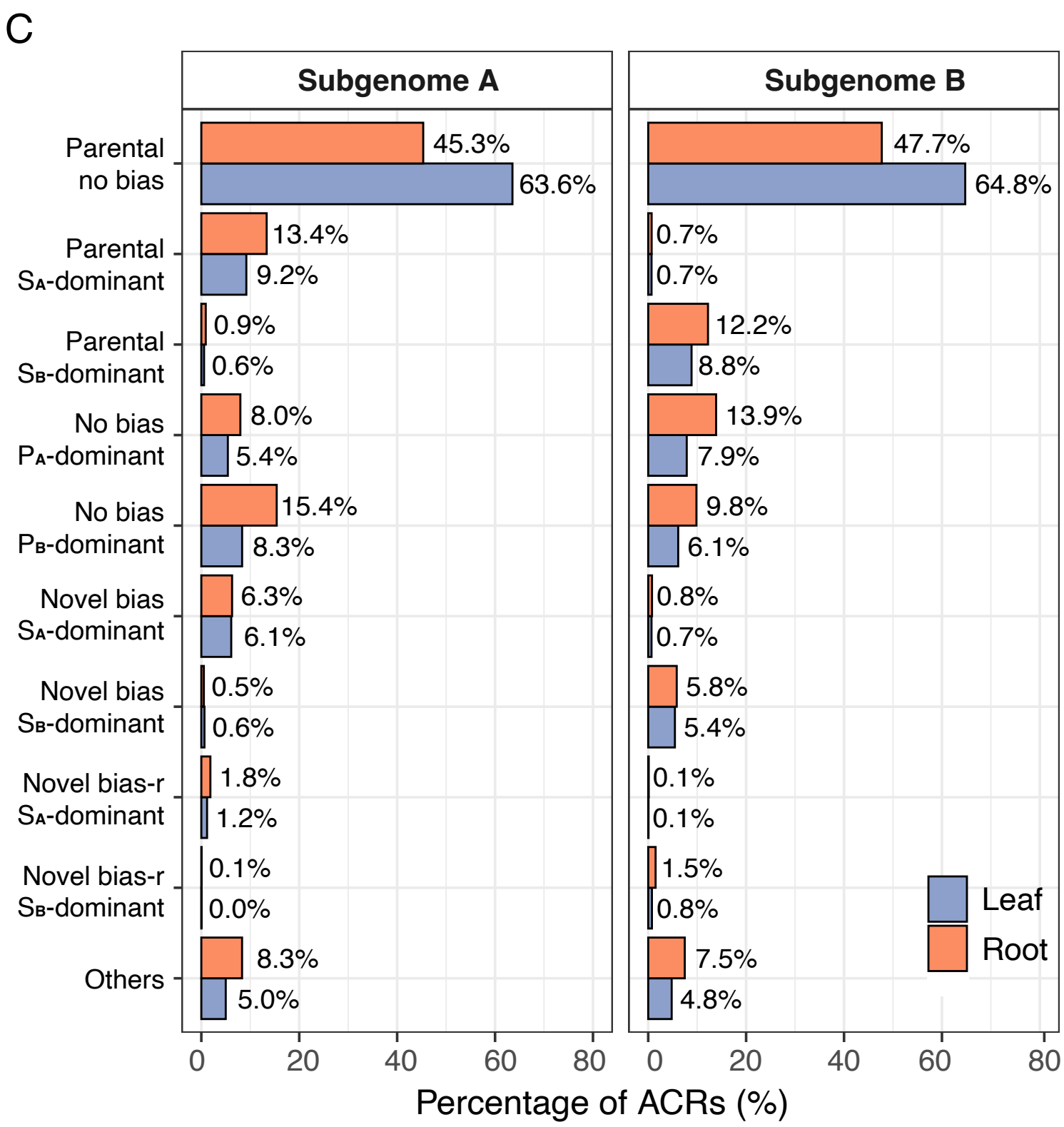
989 **(A)** Principal component analyses (PCA) of normalized expression counts for homoeologous
990 genes across four genome types: two subgenomes and their corresponding diploid progenitors
991 (top) and across two tissue types: leaf and root (bottom). **(B)** Distribution of homeolog
992 expression bias categories across leaf and root tissues. Biased expression categories were
993 labelled with red font. **(C)** Heatmap showing enrichment patterns between expression types
994 (no-bias, novel bias and parental-legacy bias) and ACR evolutionary types including peanut-
995 conserved (no bias), peanut-conserved (novel bias) peanut-conserved (parental-legacy bias),
996 lineage-specific, polyploid-specific and subgenome-specific ACRs. Absolute Z-scores greater than
997 2 are indicated.
998

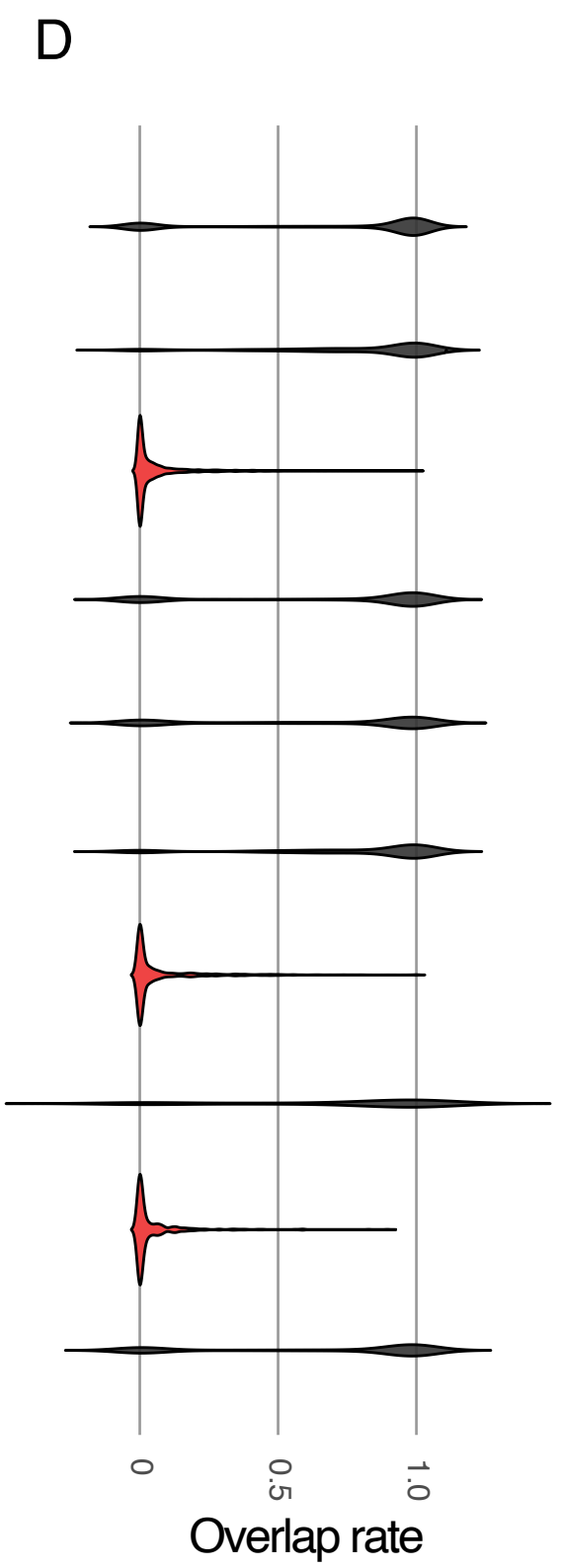

A
broad ACR
leaf-specific ACR
[0-150]
ATAC-seq (leaf)
ATAC-seq (root)
root-specific ACR
Gene
Ah.02g359400
Ah.01g107700
B
Broad
Root-specific
Leaf-specific
65%
21%
14%
67%
18%
16%
0
25,000
50,000
75,000
ACR number
C
Broad
Leaf-specific
Root-specific
34%
27%
39%
34%
39%
27%
15%
50%
35%
42%
29%
29%
42%
41%
18%
17%
53%
29%
Genic ACR
Proximal ACR
Distal ACR

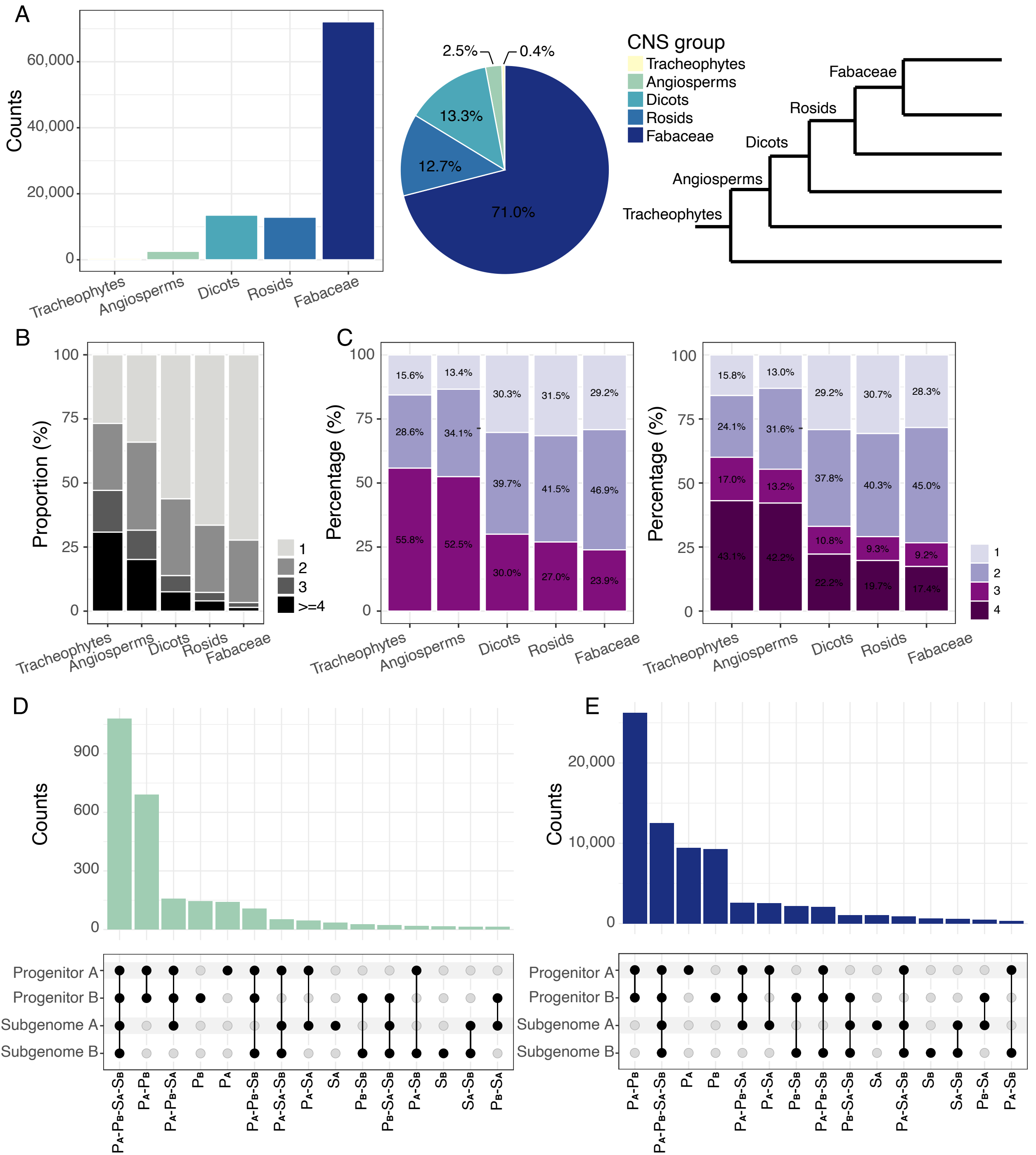

A
Progenitor A
Subgenome A
Progenitor B
Subgenome B
Peanut-conserved
Potential peanut-conserved
Lineage-specific
Polyploidy-specific
Subgenome-specific
B
Threshold (50%)
Threshold (80%)
C
100%
75%
50%
25%
0%
Lineage-specific
Peanut-conserved
Potential peanut-conserved
Polyploid-specific
Subgenome-specific
D
Mean mismatch rate
0.08
0.04
0.00
Mean BLASTed ratio
1.1
1.0
0.9
0.8
Lineage-specific
Peanut-conserved
Potential peanut-conserved
Polyploid-specific
E
H3K4me3
H3K56ac
H3K36me3
Conserved
Divergent
Ratio
0.3
0.2
0.1
PA
PB
SA
SB
Leaf Root Leaf Root
SubgenomeA SubgenomeB

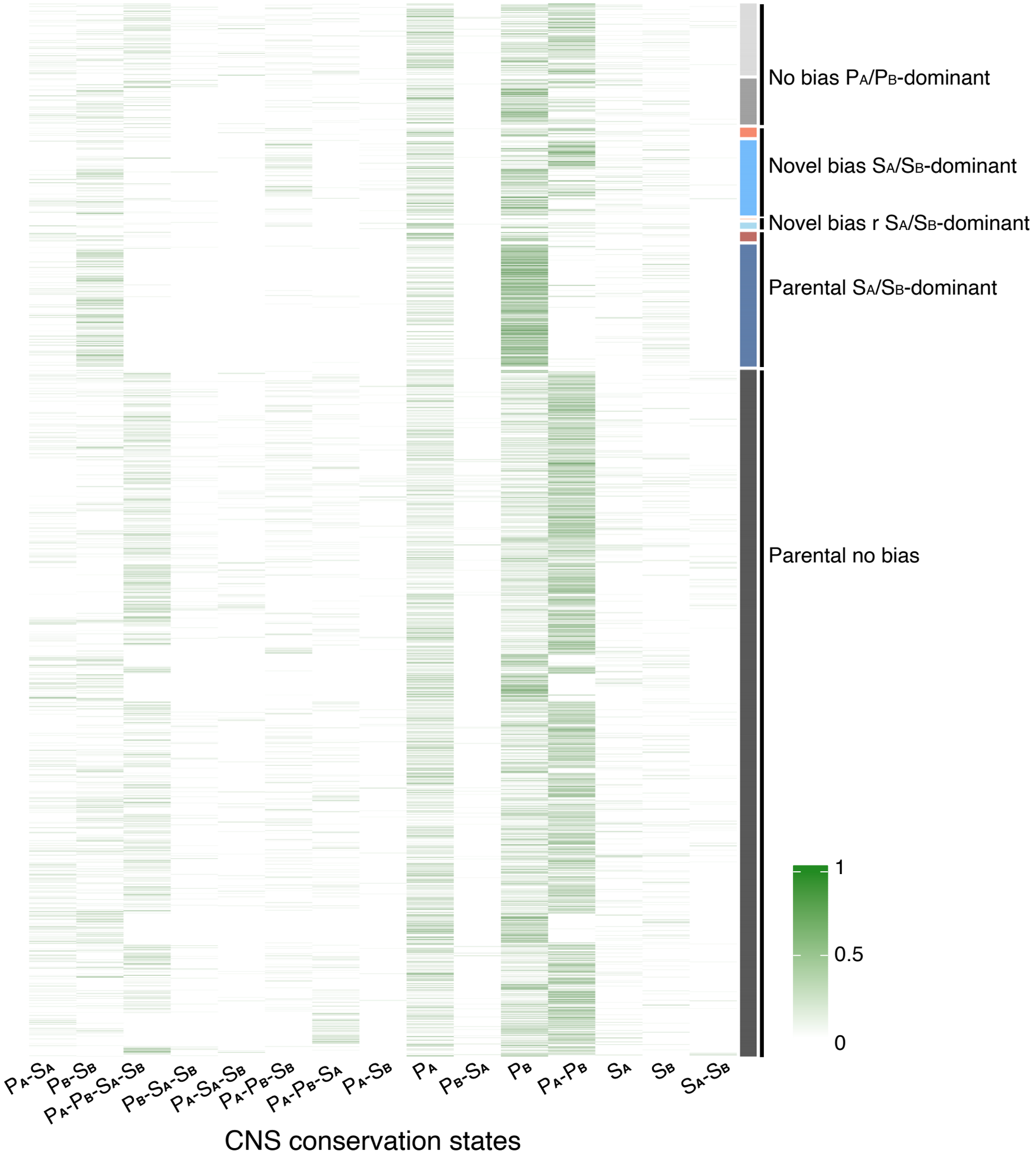

A
PC2
PC1
50
0
−50
−100
−100
−50
0
50
100
Progenitor A
Subgenome A
Progenitor B
Subgenome B
Leaf
Root

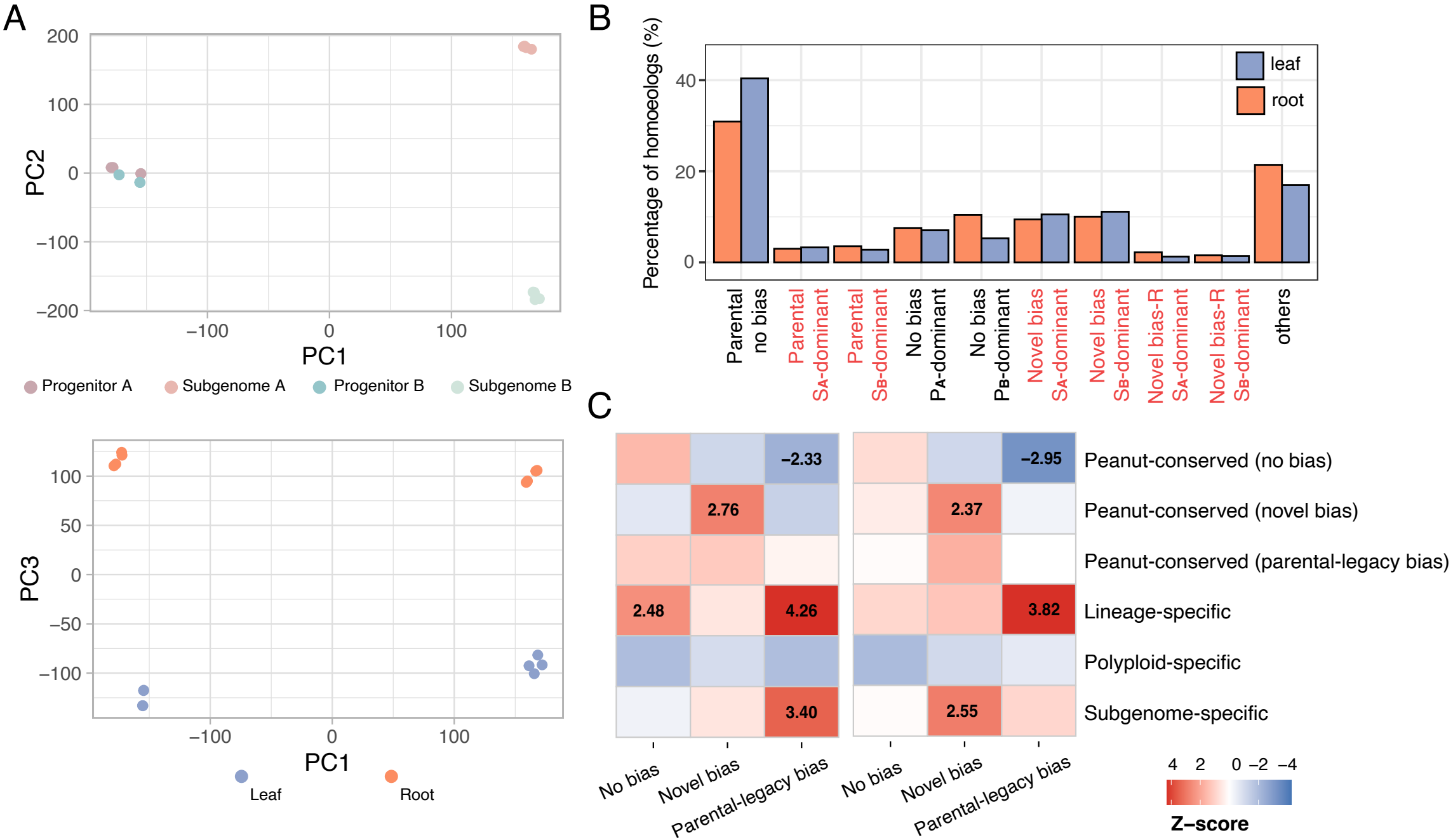

B
Progenitor A
Progenitor B
Subgenome A
Subgenome B
Parental condition
No bias in progeny
Novel bias in progeny


C
Subgenome A
Subgenome B
Parental no bias
45.3%
63.6%
47.7%
64.8%
Parental SA-dominant
13.4%
9.2%
0.7%
0.7%
Parental SB-dominant
0.9%
0.6%
12.2%
8.8%
No bias PA-dominant
8.0%
5.4%
13.9%
7.9%
No bias PB-dominant
15.4%
8.3%
9.8%
6.1%
Novel bias SA-dominant
6.3%
6.1%
0.8%
0.7%
Novel bias SB-dominant
0.5%
0.6%
5.8%
5.4%
Novel bias-r SA-dominant
1.8%
1.2%
0.1%
0.1%
Novel bias-r SB-dominant
0.1%
0.0%
1.5%
0.8%
Others
8.3%
5.0%
7.5%
4.8%
Leaf
Root
0
20
40
60
80
Percentage of ACRs (%)


D
0
0.5
1.0
Overlap rate


A
Counts
60,000
40,000
20,000
0
Tracheophytes
Angiosperms
Dicots
Rosids
Fabaceae
2.5%
0.4%
13.3%
12.7%
71.0%
CNS group
Tracheophytes
Angiosperms
Dicots
Rosids
Fabaceae
B
Proportion (%)
100
75
50
25
0
1
2
3
>=4
C
Percentage (%)
15.6%
13.4%
30.3%
31.5%
29.2%
28.6%
34.1%
39.7%
41.5%
46.9%
55.8%
52.5%
30.0%
27.0%
23.9%
15.8%
13.0%
29.2%
30.7%
28.3%
24.1%
31.6%
37.8%
40.3%
45.0%
17.0%
13.2%
10.8%
9.3%
9.2%
43.1%
42.2%
22.2%
19.7%
17.4%
1
2
3
4
D
Counts
900
600
300
0
Progenitor A
Progenitor B
Subgenome A
Subgenome B
PA-PB-SA-SB
PA-PB
PA-PB-SA
PB
PA
PA-PB-SB
PA-SA-SB
PA-SA
SA
PB-SB
PB-SA-SB
PA-SB
SB
SA-SB
PB-SA
E
Counts
20,000
10,000
0
Progenitor A
Progenitor B
Subgenome A
Subgenome B
PA-PB
PA-PB-SA-SB
PA
PB
PA-PB-SA
PA-SA
PB-SB
PA-PB-SB
PB-SA-SB
SA
PA-SA-SB
SB
SA-SB
PB-SA
PA-SB


No bias PA/PB-dominant
Novel bias SA/SB-dominant
Novel bias r SA/SB-dominant
Parental SA/SB-dominant
Parental no bias
1
0.5
0
PA-SA
PB-SB
PA-PB-SA-SB
PB-SA-SB
PA-SA-SB
PA-PB-SB
PA-PB-SA
PA-SB
PA
PB-SA
PB
PA-PB
SA
SB
SA-SB
CNS conservation states


A
PC2
PC1
Progenitor A
Subgenome A
Progenitor B
Subgenome B
B
Percentage of homoeologs (%)
leaf
root
Parental no bias
Parental SA-dominant
Parental SB-dominant
No bias PA-dominant
No bias PB-dominant
Novel bias SA-dominant
Novel bias SB-dominant
Novel bias-R SA-dominant
Novel bias-R SB-dominant
others
C
PC3
PC1
Leaf
Root
Peanut-conserved (no bias)
Peanut-conserved (novel bias)
Peanut-conserved (parental-legacy bias)
Lineage-specific
Polyploid-specific
Subgenome-specific
No bias
Novel bias
Parental-legacy bias
Z–score