Evolutionary diversity of cell-type-specific expression and stress response in Brassicaceae roots

Guannan Wang[1, 2, 3, 8], Kook Hui Ryu[4, 8], Andrea Dinneny[2, 3, 8], Jiyoung Lee[5], Dong-Ha Oh[1], Prashanth Ramachandran[2, 3], Marina Oliva[6], Ryan Lister[6, 7], José R. Dinneny[2, 3*], John Schiefelbein[4*], Maheshi Dassanayake[1*]

[1]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA

[2]Department of Biology, Stanford University, Stanford, CA, USA

[3]Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA

[4]Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI, USA

[5]Graduate Program in Genetics, Bioinformatics, and Computational Biology, Virginia Tech, Blacksburg, VA, USA

[6]ARC Centre of Excellence in Plant Energy Biology, School of Molecular Sciences, University of Western Australia, Perth, WA, Australia

[7]The Harry Perkins Institute of Medical Research, QEII Medical Centre and Centre for Medical Research, The University of Western Australia, Perth, WA, Australia

[8]These authors contributed equally to this work.

*Correspondence: dinneny@stanford.edu (J.R.D.); schiefel@umich.edu (J.S.); maheshid@lsu.edu (M.D.)

ABSTRACT

Plants are composed of diverse cell types that facilitate adaptations to the environment, yet cross-species comparisons of such response programs at single-cell resolution remain scarce. To explore this diversity, here we profiled >200,000 root cells from five Brassicaceae species, including stress-sensitive species (*Arabidopsis thaliana, Sisymbrium irio*), extremophytes (*Eutrema salsugineum, Schrenkiella parvula*), and the polyploid crop *Camelina sativa* under control, NaCl, and abscisic acid (ABA) treatments. We developed a computational pipeline to characterize the conservation and divergence of cell-type gene expression across the Brassicaceae, revealing that approximately half of previously defined Arabidopsis cell-type markers fail to maintain conserved expression in one or more non-Arabidopsis species. Therefore, we curated a refined set of pan-Brassicaceae markers and identified orthologs whose expression profiles have diverged across lineages. Using *in situ* hybridization, we mapped distinct cortex subpopulations to specific cortical layers across species, and found flavonoid biosynthesis programs preferentially localized to the inner cortex layer, linking cortex-layer specification to metabolic specialization. Cell-type contributions to stress responses differed among species and across treatments, with lineage-specific losses of responsiveness occurring less frequently but evolutionarily more favored than lineage-specific gains. In *C. sativa*, sub-genomes contributed equally to stress responses, and homeologs with divergent responses typically lacked other signatures of functional divergence. Together, this work establishes a foundational root single-cell atlas and an analytical framework for multi-species comparative transcriptomics, providing insights into how stress responses diversify across cell types, stress-sensitive to stress-adapted species, and in crop lineages.

KEYWORDS

extremophytes, oil crop, single cell transcriptomics, comparative transcriptomics, ABA, salt-stress adaptation, ortholog expression, root development, homeologs, evolution

INTRODUCTION

Environmental stress responses are complex traits that require spatiotemporal coordination of gene networks functioning in diverse assemblages of cell types. Plant roots, the primary interface with soil, are a mixture of functionally and transcriptionally diverse cell types, each perceiving and responding to environmental stresses in a development-dependent way[1–3]. Investigations of plant environmental responses at the cell-type level have been limited to model plants like *Arabidopsis thaliana*[4,5], and thus provide only partial insights into the diverse physiological strategies employed by plants. This limitation becomes more evident when considering the rewired and novel stress tolerance mechanisms employed by naturally stress-adapted species (extremophytes) that exhibit a remarkable ability to optimize growth and survival under challenging environments[6]. However, such comparisons are complicated in distantly related species, where limited one-to-one orthology constrains evolutionary inference at the level of individual genes. Therefore, establishment of a comparative approach within a model clade of related species not only enhances our understanding of the diversity of stress adaptations but also uncovers evolutionary innovations that confer stress resilience.

In this study, we have selected ecologically diverse species within the Brassicaceae family that include the extremophyte models, *Eutrema salsugineum* and *Schrenkiella parvula,* the more stress-sensitive models, *Arabidopsis thaliana* (hereinafter referred to as Arabidopsis) and *Sisymbrium irio* (a close relative of the extremophyte models that exhibits better salt stress tolerance compared to Arabidopsis[7]), and the polyploid oil crop, *Camelina sativa*. Previous studies reported divergent physiological responses among these species not only to high salinity, but also to the stress hormone, abscisic acid (ABA)[8–10]. Moreover, it has been demonstrated that stress conditions can elicit distinct responses across different root cell types, and these cell-type specific responses can vary between species[1,2,11,12]. However, how individual cell types contribute to the diversity of stress responses remains largely unknown. We generated primary root transcriptome atlases for the selected species at single-cell resolution under stress-neutral as well as ABA- and salt-treated conditions, totaling >200,000 cells across five species (>174,000 cells across four diploid species plus >27,000 cells in *C. sativa*). This experimental approach allowed us to compare recently diverged lineages and to examine responses within a defined phylogenetic group with a shared common ancestor ~33 Myo[13]. We developed a comparative framework and methodology for cross-species cell-type annotation and single-cell evolutionary transcriptomics that can be readily transferred to other multi-species comparisons. Using this framework, we (i) quantify the limited conservation of canonical Arabidopsis cell-type markers across the Brassicaeae and define a refined set of conserved markers; (ii) identify significant changes in ortholog expression profile on different phylogenetic branches; (iii) spatially map cortex subpopulations to defined cortex layers across species and provide evidence for cortex functional specialization; (iv) resolve how distinct cell types contribute to salt versus ABA responses across lineages and identify patterns of evolutionary turnover in responsiveness; and (v) evaluate the extent to which subgenomes and homeologs diversify stress responses in the polyploid crop *C. sativa*. Together, this work establishes a foundational root cell atlas to explore cell-specific, species-specific, or stress-dependent mechanisms across ecologically diverse but phylogenetically close Brassicaceae model plants.

RESULTS

The expression of nearly half of the detected Arabidopsis cell type marker genes is not conserved across species

We generated 36,000 to 50,000 root cell transcriptomes for four diploid Brassicaceae species under control (stress-neutral), 5 µM ABA, and 100 mM NaCl treatments (n = 2 for each condition) using the 10X Genomics scRNA-seq platform to provide over 174,000 root cell transcriptomes from the four diploid species (Fig. 1 and Supplementary Fig. 1; Supplementary Movie 1). To investigate the impact of genome duplication on stress adaptation, we additionally generated a comparable dataset of over 27,000 root cell transcriptomes for an allohexaploid Brassicaceae species using an identical experimental design. The percentage of orthologs between species generally decreases with increasing species divergence (Supplementary Fig. 2), which limits comparative studies investigating large ortholog gene networks. We therefore selected more closely-related species (diverged max. ~33Mya) of Arabidopsis[13] from the Brassciceae family for our comparison, including the diploid *E. salsugineum*, *S. irio*, *S. parvula*,

112 and hexaploid *C. sativa*. This approach allowed us to identify 15,198 unambiguous 1-to-1
113 orthologous genes across the diploid species and the Arabidopsis orthologs for over 90% of
114 genes in *C. sativa* (details in Methods). Given the complex 1-to-3 orthology relationship inherent
115 to hexaploidy, we intentionally excluded *C. sativa* from 1-to-1 ortholog based cross-species
116 comparisons (Figs. 1-6) and analyzed this species separately (Fig. 7) to preserve sub-genome
117 resolution. Regardless of the species and conditions, we consistently generated an average of
118 35,077 reads for each cell, resulting in the detection of a median of 1,959 unique transcripts and
119 1,077 genes expressed per cell (Supplementary Data 1 and Supplementary Fig. 3). We used ABA
120 as a more generic abiotic stress stimulus and NaCl as a more specific stimulus. We selected
121 treatment strength and duration based on our prior work showing that ABA (μM, 3h) and NaCl
122 (100-150 mM, 24h) trigger strong root transcriptomic responses and physiological changes
123 across species, including extremophytes, without causing immediate tissue necrosis in sensitive
124 plants like Arabidopsis[9,10,12,14]. To facilitate the utility of this multi-species dataset as a
125 community resource, we generated an accompanying interactive web browser to enable
126 exploration of this atlas at https://plantbiology.shinyapps.io/brassicaceaeatlas. Preprocessed data
127 matrices partitioned by cell types are provided in the searchable browser for convenient
128 reanalysis, as well.

129 Root cell-type marker genes have been established for Arabidopsis[15–21], though their
130 applicability as markers for root cell types in related species has been untested. We developed a
131 two-step strategy to annotate cell types for Arabidopsis and non-Arabidopsis cells in our dataset
132 (Supplementary Fig. 4; Methods). We used Arabidopsis cells from the control condition, and
133 annotated these individual cells based on the collective outcome of three independent methods:
134 correlation-based annotation (where individual cells were annotated based on their
135 transcriptomic similarity to reference cell type transcriptomes), marker-based annotation (where
136 individual cells were annotated based on the preferential expression of known cell type markers),
137 and integration-based annotation (where individual cells were annotated based on their mapping
138 to reference single-cell datasets) (Supplementary Fig. 4 - Step 1). This provided us with a
139 reference atlas to identify cell types for the remaining species. We employed three prediction
140 algorithms (TransferData in Seurat[22], Symphony[23], and scArches[24]) to transfer Arabidopsis
141 annotations to individual cells in the other species based on 15,198 unambiguous 1-to-1
142 orthologous genes identified (Supplementary Fig. 4 - Step 2, details in Methods). We
143 additionally annotated our dataset with differential developmental zones using a similar
144 approach. This resulted in the identification of ten major root cell types and three developmental
145 zones for all four species from all conditions in our dataset (visualized on the integrated UMAP
146 space, Figs. 1a-1c, Supplementary Fig. 5 and 6). We did not detect a recognizable cell population
147 for quiescent center (QC) cells in our dataset (Supplementary Fig. 7), likely due to under-
148 sampling of this rare cell type. Our overall cell-type identification indicates that these diverse
149 Brassicaceae species do not differ in the presence of major root cell types.

150 The expression conservation of known Arabidopsis cell-type marker genes was examined in the
151 other target species (Fig. 1d and Supplementary Fig. 8). Out of 461 previously established
152 Arabidopsis cell-type markers, we found 1-to-1 orthologs for 254 of them, 241 of which were
153 representative marker genes for the ten major root cell types described above (Supplementary

Data 2). Only 144 of these Arabidopsis markers had expression levels above the threshold of detection in our Arabidopsis data (Fig. 1e). Among these Arabidopsis markers, while many retained robust expression and cell-type specificity across all species, 48% (69/144) of markers lost expression specificity in one or more species (Figs. 1d and 1e, Supplementary Fig. 8 and 9). Among the non-conserved Arabidopsis markers, lost conservation can be attributed in most cases to low detectable expression in the expected target cell types in one or more non-Arabidopsis species (Supplementary Fig. 9b). The lowest rate of conservation for Arabidopsis cell-type markers was observed in *E. salsugineum* (Fig. 1e and Supplementary Fig. 8) despite the data quality for this species being comparable to that of other species (Supplementary Data 1). The cell-type markers for vascular tissues were generally highly conserved across species, while markers from outer cell layers of the root, especially the root cap and epidermal cells, were least conserved (Supplementary Fig. 8). Similar patterns of conservation were also observed when comparing tissue-specific gene expression programs between environmental conditions in Arabidopsis[1].

To obtain a panel of pan-Brassicaceae cell type markers, we identified orthologous genes exhibiting highly conserved expression for each of the ten cell types (Supplementary Fig. 10; Supplementary Data 3). To test the accuracy of this panel, we optimized a whole-mount hybridization chain reaction (HCR)-based *in situ* hybridization method[25] for *in vivo* mapping of the selected markers. *In situ* expression patterns matched the expected cell-type specificity based on the scRNA-seq data (Fig. 2 and Supplementary Fig. 11). This provided independent support for the cell-type assignments in the target species and shows that these orthologous genes represent broadly conserved cell-type markers.

Altered cell-type expression specificity drives lineage-specific molecular phenotypes

We observed that orthologs that retained similar cell type specificity across all four species accounted for only a small proportion (~ 3% of the 1-to-1 orthologs with all pairwise Pearson correlations greater than 0.9) of the entire ortholog sets (Fig. 3a and Supplementary Fig. 12). The overall correlations of cell-type expression among orthologs were modest even between each pair of species (Fig. 3a). To determine the patterns of divergence in cell-type expression between species under stress-neutral conditions, we compared 1-to-1 ortholog expression across the four species using a phylogenetic framework, and detected 350 orthologs showing diverged expression profiles in one phylogenetic branch compared to others or between extremophytes and non-extremophytes (Fig. 3b; Supplementary Data 4). For example, there were 78 orthologs whose expression pattern in Arabidopsis (lineage I) differed from that of orthologs in the other three (lineage-II) species. We also found 43 orthologs whose cell-type expression differentiate extremophyte and non-extremophyte species (Fig. 3b), independent of their phylogenetic relationships.

We used HCR to examine two of top-ranked diverged cell-type expression profile changes associated with the extremophytes: restricted expression of *NOD26-like intrinsic protein 5;1* (*NIP5;1*) in *S. parvula*, and extremophyte-shared expression of *glutathione S-transferase U25* (*GSTU25*) (Figs. 3c and 3d). As both *NIP5;1* and *GSTU25* are members of larger gene families, to ensure the detected expression changes across species is not due to the misassignment of

195 orthologs from the same gene family, we identified all *NIP*s and *GSTU*s among these species and
196 extracted their protein sequences for phylogeny construction. As expected, *NIP5;1* and *GSTU25*
197 orthologs used in our analyses formed their own branches separated from all other *NIP*s and
198 *GSTU*s, respectively, confirming the assignment of orthologs (Supplementary Fig. 13). In the
199 *GSTU25* branch, we noticed the presence of an additional gene from *S. parvula*, *Sp1g15325*,
200 which is a non-reciprocal-best-match for *GSTU25*s from other species (Supplementary Fig. 13).
201 It exhibited a *SpGSTU25*-like cell type expression profile that was shared with *E. salsugineum*
202 but not with the non-extremophytes (Supplementary Fig. 14).

203 Both scRNA-seq data and HCR results showed that the gene *NIP5;1*, which encodes a plasma
204 membrane-localized aquaporin, is predominantly expressed in the outer cell layers, from
205 epidermal to endodermal cells, in Arabidopsis, *S. irio*, and *E. salsugineum,* while its expression
206 is restricted to differentiation zone endodermal cells in *S. parvula* (Figs. 3c and 3e). Arabidopsis
207 AtNIP5;1 serves as a channel responsible for transporting boron into roots[26]. The restricted
208 expression of *NIP5;1* in *S. parvula* may reduce boron uptake into the vascular stream. *S. parvula*
209 is native to boron-rich soils in Anatolia, Turkey and can tolerate high boron levels[14] while the
210 other three species are not associated with exceptionally high boron habitats. Previous studies
211 have demonstrated that *S. parvula* achieves boron toxicity tolerance mainly by maintaining low
212 intracellular boron levels[27], which is consistent with the more restricted expression of *NIP5;1* in
213 this species.

214 Our scRNA-seq data indicate that *GSTU25*, encoding a plant-specific Tau class glutathione S-
215 transferase, is preferentially expressed in the root cap of non-extremophytes (Fig. 3d). In the
216 extremophytes, its expression becomes enriched in the endodermal cells of the meristematic and
217 early elongation zones (Fig. 3d). HCR experiments confirmed the absence of endodermal
218 expression of *GSTU25* in the non-extremophytes (Fig. 3f). However, our HCR observations did
219 not fully align with the scRNA-seq data, particularly for *E. salsugineum* where *GSTU25*
220 expression was also detected in epidermal and cortex cells (Figs. 3d and 3f). This discrepancy
221 may be due to the technical differences between the two complementary RNA detection
222 methods. Specifically, because of its signal amplification mechanism and better cell resolution,
223 HCR was able to detect the expression of *GSTU25* in the cell population that might have been
224 lost or under-sampled in scRNA-seq. Glutathione S-transferases protect cells from oxidative
225 damage by utilizing glutathione to scavenge reactive oxygen species (ROS)[28]. *GSTU25* has been
226 reported to be induced under oxidative, salt, and heat stresses[29,30]. The convergent endodermal
227 expression of *GSTU25* in the two extremophytes might suggest a role of *GSTU25* in ROS
228 scavenging in endodermal cells, which may contribute to salt tolerance in the two
229 extremophytes.

230 **Transcriptomic specialization of cortex cell subpopulations supports lineage-specific root**
231 **anatomy**

232 Anatomical analysis of *S. irio* and *S. parvula* reveals that both species possess two cortical cell
233 layers, whereas Arabidopsis and *E. salsugineum* possess a single cortical cell layer at the early
234 developmental stage of seedlings (Supplementary Fig. 15a). Cortex cells from the single cortical
235 layer in Arabidopsis have been reported to contain two transcriptionally distinct

236 subpopulations[31]. Agreeing with this observation, UMAP-based analysis of cortex cells in our
237 dataset identified two major cortex subpopulations for Arabidopsis and *E. salsugineum*.
238 Contrastingly, three cortex subpopulations were found in *S. parvula* and *S. irio* (Supplementary
239 Fig. 15b). We hypothesized that the extra cortex cell population identified in *S. irio* and *S,*
240 *parvula* is derived from the additional cortex layer. If supported, we would expect that the extra
241 cortex cell population should be absent or underrepresented in Arabidopsis and *E. salsugineum*.
242 To test this, we selected the cortex subpopulations from *S. irio*, which had the highest separation
243 of clusters, as the reference and categorized the three subpopulations from this species into
244 "left", "middle", and "right" subgroups (Fig. 4a). Cortex cells from the other three species were
245 mapped to the reference *S. irio* cortex subpopulations based on 1-to-1 orthologs (Fig. 4b). While
246 some representation of the three cortex subpopulations could be detected in all species (Figs. 4b
247 and 4c), the middle population was highly depleted in Arabidopsis and *E. salsugineum*, but not in
248 *S. parvula* (Fig. 4c). To test whether the "middle" cortex subpopulation represented the extra
249 cortex layer found in *S. irio* and *S. parvula*, we identified marker genes that were preferentially
250 expressed in the "middle" population (Supplementary Fig. 16a; Supplementary Data 5) and
251 examined their spatial expression in roots using HCR. The *in situ* hybridization results showed
252 that *Si-s1384-01650*, which was specifically expressed in the middle population (Fig. 4d, upper
253 panel), distinctly marked the inner cortex layer in *S. irio* (Fig. 4d, lower panel), in support of our
254 hypothesis. Similarly, the ortholog (in the ortholog group OG14187) of *Si-s1384-01650* showed
255 expression predominantly in the inner cortex layer of *S. parvul*a, while in *E. salsugineum* and
256 Arabidopsis, expression was detected in all cells of the single cortex layer (Fig. 4e). These
257 results suggest that the middle subpopulation is derived from the inner cortex layer, while the
258 other two populations originate from the outer layer. To characterize the transcriptomic
259 signatures of the inner and outer cortex layers, we identified genes that were preferentially
260 expressed in either of the layers (Supplementary Fig. 17). Both layers highly expressed genes
261 involved in water transport and environmental sensing (Supplementary Fig. 17). Genes
262 preferentially expressed in the inner cortex are mostly associated with secondary metabolism
263 (particularly flavonoid biosynthesis), cell-to-cell communication via plasmodesmata, and
264 jasmonic acid and ABA signaling. In contrast, the outer cortex prioritized hypoxia sensing,
265 thermal stress response and gibberellin signaling pathway gene functions (Supplementary Fig.
266 17).

267 To further investigate the transcriptomic differences between the inner and outer cortex layers in
268 *S. irio* under control conditions, we compared the transcriptomes between the inferred inner
269 (middle subpopulation) and outer cortex layers (left and right subpopulations, Supplementary
270 Fig. 16). This identified several cellular processes (among 155 genes that were differentially
271 expressed between the two layers) that were differently partitioned between the two layers,
272 including flavonoid biosynthesis (Supplementary Fig. 16) which has been previously shown to
273 be preferentially localized to cortex cells[32]. We detected a strong expression bias towards the
274 inner layer in *S. irio* for about 69% of the genes involved in flavonoid biosynthesis, including
275 CHALCONE SYNTHASE (CHS) (Fig. 4f, left panel), which catalyzes the first committed step
276 in this biosynthetic pathway. Intriguingly, this inner cortex layer-bias is also conserved in *S.*
277 *parvula* cortex cells (Fig. 4f, right panel), suggesting that this transcriptional bias preceded the
278 divergence of this subclade of species. Similar biases in flavonoid biosynthetic gene expression

were not observed in other species with more than one layer of cortex (Supplementary Fig. 18), including rice[33] and Medicago[34], suggesting that the partitioning of flavonoid biosynthesis is unlikely to be a conserved signature of cortex functional diversification. Flavonoids are known to function as antioxidants, modulators of reactive oxygen species and hormone transport and signaling, and have been implicated in plant tolerance to environmental stresses[35,36]. The spatial restriction of flavonoid biosynthesis to the inner cortex therefore suggests a lineage-specific partitioning of a growth and stress-relevant metabolic pathway within the root. Together these data demonstrate transcriptional divergence in the cortex cell populations, corresponding with the increased anatomical complexity observed in *S. irio* and *S. parvula* roots. Our results suggest that the presence of two cortex layers may have enabled the inner layer to specialize in flavonoid metabolism.

ABA and NaCl induce distinct cell-type-specific responses

To obtain an overview of cell-type transcriptional diversity across species and in response to stress inducers, we performed a principal component analysis (PCA) using the 15,198 1-to-1 orthologous genes from pseudo-bulk cell-type transcriptomes (Fig. 5a). Although the transcriptomes of all cell types clustered according to their respective species, those originating from the two extremophytes formed a distinct subgroup that was separate from the transcriptomes of other non-extremophytes. Since *S. parvula* and *S. irio* are more closely related phylogenetically than are the two extremophytes, the observed similarities between extremophytes suggests that these shared features arose independently of phylogenetic relatedness. Within each species, outer root cell layers (root cap, epidermal, cortex, and endodermal cells) formed a distinct cluster from the inner cell layers (pericycle, procambium, phloem, and xylem cells), regardless of treatments. We next tested if these observed differences could be a consequence of protoplasting using published single-nucleus RNA-seq for Arabidopsis[37] where cells were not subjected to a protoplasting step. The two groups, inner and outer cell types, remained as two distinct groups (Supplementary Fig. 19). Therefore, the transcriptomic divergence between the inner and outer cell layers suggests that these two zones of the root operate within distinct signaling contexts and exhibit a more evolutionary conserved and robust developmental trajectory across species.

To explore which genes, processes, or cell states drive the overall distinct stress responses between inner and outer cell layers, we next identified differentially expressed genes (DEGs) on a per cell type, species, and treatment basis (details in Methods; Supplementary Data 6). To determine the stress response in each developmental zone, we applied the same approach to identify DEGs in all zones (Supplementary Data 6). We selected three ortholog groups exhibiting species-specific transcriptional responses to explore using HCR. For example, in response to ABA, members of orthologous group OG04736 (AT3G15670), which encode a late embryogenesis abundant protein (LEA), showed a ubiquitous increase in expression across tissues, which was largely captured by HCR (Supplementary Fig. 20). HCR also confirmed expression changes in another ortholog group, OG03053 (AT2G37870), which exhibited the highest expression in stele cells compared to all other cell types and was further induced by ABA (Supplementary Fig. 21). In another example, the expression of the *tonoplast intrinsic protein*

320 *2;3* (*TIP2;3*; AT5G47450; OG10873) was consistently reduced in multiple cell types across
321 species with exposure to NaCl. The HCR signal for *TIP2;3*s displayed a similar trend in
322 expression across all species (Supplementary Fig. 22). These data support the overall trends
323 observed with scRNA-seq and further attest to the fidelity of the cell-type transcriptional profiles
324 generated through single cell protoplasting and scRNA-seq analysis.

325 Previous studies have identified Arabidopsis as the most sensitive of the four species, exhibiting
326 the strongest transcriptional response to ABA- and NaCl-induced stress[9,10]. Our results support
327 this finding, showing Arabidopsis has the highest number of differentially expressed genes
328 (DEGs), ranging from 2 to 10 times more than the other species (Supplementary Fig. 23). The
329 number of DEGs also differed among different developmental zones across species with a
330 substantial portion of them being specific to individual zones (Supplementary Fig. 24). To
331 compare the responses to ABA and NaCl treatments, we performed a direct comparison of stress
332 responses by identifying genes that showed shared or distinct responses to ABA and NaCl
333 (Supplementary Fig. 25). This quantitative assessment revealed that genes responsive to both
334 stimuli typically accounted for less than 50% of the total responsive genes, with a notable
335 exception for Arabidopsis where up to 78% of the ABA-responsive genes were shared with NaCl
336 stress response (Supplementary Fig. 25). While these results agree with previously reported
337 overlap between these two treatments[9], it also suggests that the majority of the transcriptomic
338 remodeling is indeed stress-specific, and salt stress activates not only ABA-mediated responses
339 but also a large additional set of ABA-independent pathways.

340 Quantification of cell type responsiveness using the number of DEGs can be confounded by
341 factors such as the number of cells per cell type and different effect sizes of genes[38]. To assess
342 cell-type response signals isolated from internal noise or other confounding factors, we
343 implemented Augur[38], a method to rank cell types based on the magnitude of their response in
344 single-cell data. Procambium cells showed the strongest response to ABA in Arabidopsis and *E.*
345 *salsugineum,* whereas columella cells were the most responsive in *S. irio* and *S. parvula* (Fig.
346 5b). The species-specific cell-type responses further intensified when plants were treated with
347 NaCl. In Arabidopsis and *E. salsugineum*, inner cell layers were the most responsive cell types,
348 whereas in *S. irio* and *S. parvula* the root cap was most responsive (Fig. 5b). While the two
349 extremophytes' transcriptomes cluster closer to each other (Fig. 5a), the overall normalized
350 transcriptional responses to stress in each species indicated that *S. parvula* and *S. irio* were more
351 similar, aligning with their closer phylogenetic relationship (Fig. 5c).

352 We examined the functional processes enriched in DEGs to identify the most predominant
353 cellular processes associated with stress responses (Fig. 5d; Supplementary Data 7). Response to
354 stress was among the most commonly shared processes enriched in DEGs induced by ABA and
355 those repressed by NaCl across most cell types in all four species. RNA metabolism and
356 ribosome biogenesis were found exclusively enriched among repressed genes in response to
357 ABA. Additionally, these processes exhibited the greatest variation among cell-type specific
358 responses across species, being primarily enriched in endodermal and stele cells in Arabidopsis
359 and *S. irio*, but in atrichoblast cells in *S. parvula* (Fig. 5d). On the contrary, these same processes
360 were enriched only among induced DEGs in response to NaCl across species. Moreover, these

361 over-represented functions were primarily found in outer cell types in Arabidopsis, but restricted
362 to only pericycle cells in *E. salsugineum* and epidermal cells in *S. parvula* (Fig. 5d). These
363 results indicate that although stress signaling involves similar biological functions in each
364 species, the key differences lie in the cell-specific patterns of these responses.

365 The observed divergence in stress responses among species seemed to primarily occur between
366 cell types. Therefore, we computed alterations in transcriptomic distinctiveness between pairs of
367 species for each cell type in response to ABA and NaCl treatments. This allowed us to pinpoint
368 the cell types that exhibited the most pronounced divergence in their responses. We observed
369 overall larger transcriptomic distinctiveness between species in response to 100 mM NaCl
370 compared to that of 5 μM ABA (Fig. 5e, Supplementary Fig. 26). The columella cells exhibited
371 the highest transcriptomic divergence compared to other cell types under stress-neutral
372 conditions (Fig. 5e and Supplementary Fig. 27), consistent with a previous report for grass
373 species[39]. ABA and NaCl treatments did not dramatically alter the high level of columella
374 transcriptomic divergence between species (Supplementary Fig. 27). In contrast, trichoblast cells
375 showed the most pronounced shifts for between-species transcriptomic distinctiveness upon
376 NaCl treatment where transcriptomic divergence increased between Arabidopsis and lineage II
377 species but decreased among lineage II species. Cortex and atrichoblast transcriptomes
378 significantly increased their similarity following NaCl treatment only between the two
379 extremophytes (Fig. 5f, quantified as NaCl-induced reduction in transcriptomic distinctiveness
380 between the extremophytes in Supplementary Fig. 28), indicating convergent salt-induced
381 transcriptomic adjustments in these two cell types for the extremophytes. Procambium showed
382 the most prevalent increases in transcriptomic divergence between species, particularly in
383 response to 5 μM ABA treatment, whereas most other cell types remained transcriptomically
384 similar between species (Fig. 5f, and Supplementary Figs. 27 and 28). Since procambium
385 orchestrates the development of vascular tissues and is sensitive to stress and hormonal
386 signaling, the pronounced divergence observed in procambium suggests that stress-responsive
387 transcriptional programs may be preferentially rewired in tissues coordinating growth and
388 transport rather than being restricted to terminally differentiated cell types.

389 **Phylogenetic modeling reveals asymmetric evolutionary rewiring of stress responses across**
390 **species**

391 We aimed to investigate the occurrences of ortholog response shifts between stress-sensitive
392 differential expression (DEGs) and stress-insensitive expression (non-DEGs), examining
393 whether such changes correlated to phylogenetic relationships. For this analysis, we considered
394 all ortholog groups that had at least one DEG responsive to ABA or salt in at least one cell type.
395 There were more DEGs responsive to salt than to ABA in all species (Supplementary Fig. 23).
396 Consequently, the responsive ortholog groups were more frequent for salt than for ABA (a total
397 of 6,848 and 30,133 ortholog groups from all cell types for ABA and NaCl response,
398 respectively) (Supplementary Fig. 29).

399 To determine whether the observed expression shifts were greater than expected by chance, we
400 developed a bootstrapping approach that preserved the original data structure.This allowed us to
401 generate an expected distribution and calculate the *p*-values for each pattern (see Methods). A

402 gain of response was recorded when an ortholog switched its expression from a non-DEG to a
403 DEG, whereas a loss of response was recorded when an ortholog switched from a DEG to a non-
404 DEG within one lineage, relative to all other orthologs in the group. Lineage-specific gains in
405 stress response were consistently observed less than expected and exhibited a wide distribution
406 across cell types and species (Fig. 6a and Supplementary Fig. 29). Contrastingly, frequencies for
407 lineage-specific loss of stress responses exceeded expected frequencies. These occurrences were
408 confined to pericycle and procambium cells in response to ABA but were pervasive across most
409 cell types under salt stress (Fig. 6a and Supplementary Fig. 29). This distinctive pattern suggests
410 that transcriptional responses to stress in xylem and phloem cells are particularly stable across
411 species. Synapomorphic (shared derived traits from a common ancestor) and convergent
412 (independently evolved) responses were observed at frequencies greater than expected,
413 especially in comparison to divergent responses (Fig. 6a and Supplementary Fig. 29). We
414 additionally compared the response shifts among ortholog groups with gene duplication events
415 and observed nearly identical trends as the 1-to-1 orthologs **(**Supplementary Fig. 30**)**, indicating
416 the transitions between DEGs and non-DEGs are likely to be insensitive to local duplications.

417 As shown in Fig. 6b and Supplementary Fig. 29, convergent responses were found between
418 species from different lineages regardless of their phylogenetic relationship. We identified
419 orthologs exhibiting convergent stress responses shared between the two extremophytes, and
420 found 50 and 91 orthologs responding to ABA and NaCl treatments, respectively (Fig. 6b and
421 Supplementary Fig. 29). We also detected 123 and 683 orthologs that were only responsive to
422 ABA and NaCl in non-extremophytes (Fig. 6b and Supplementary Fig. 29). While the ABA-
423 responsive orthologs shared between the two non-extremophytes were mostly enriched for
424 nucleic acid metabolism, those shared by the extremophytes were mostly associated with
425 response to abiotic stress (Supplementary Data 8). When exposed to NaCl, defense response
426 pathways were strongly enriched among orthologs that were induced only in non-extremphytes
427 or those orthologs repressed only in the extremophytes (Supplementary Data 8).

428 Given that ABA-responsive element binding factors (ABFs) are key regulators of ABA
429 responsive gene expression, we next asked whether the divergent ABA and salt responses
430 observed across species could, in part, result from differences in ABF binding. To answer this,
431 we compared the differential regulation of orthologs under ABA treatment with ABF binding in
432 their promoters, as previously identified using DAP-seq[9]. We observed consistently higher
433 percentages of genes with ABF promoter binding among ABA induced genes compared to
434 repressed genes or NaCl responsive genes (Supplementary Fig. 31). Additionally, regardless of
435 species, genes with more than one ABF binding event were generally differentially regulated in
436 more cell types than those with only one ABF binding event upon ABA treatment, and to a much
437 lesser extent, NaCl treatment (Supplementary Fig. 32). Interestingly, we found that orthologs
438 with divergent ABF binding were more likely to be divergently regulated by ABA, but not by
439 NaCl (Fig. 6c and Supplementary Fig. 33). There was a substantial portion of orthologs (roughly
440 86% to 94%) showing different ABA responses even when all orthologs in the group had
441 conserved ABF binding sites in their promoters (Fig. 6c). Together, these results indicate that
442 divergence in ABF binding among orthologs can partially explain their divergent responses to

443 ABA but that the divergent salt responses likely involve more complex ABA-independent
444 regulation.

445 **Cell types contribute differentially to the stress-prepared transcriptomes in the**
446 **extremophytes**

447 Previous studies examining salt stress responses at the organ level in Arabidopsis and its
448 extremophyte relatives have suggested that extremophytes exhibit stress-prepared
449 transcriptomes, characterized by minimal overall changes to stress treatment, which may serve as
450 a key determinant of their resilience to stress[10,40,41]. We aimed to investigate whether the stress-
451 prepared state at the organ level in extremophytes resulted from a collective response of
452 individual cells exhibiting a uniform low response or if contrasting responses in different cell
453 types culminated in the absence of a net transcriptomic response captured at the organ level. We
454 examined co-expression networks of all stress-responsive orthologs across species and
455 conditions from all cell types. This resulted in 18 clusters with distinct co-expression patterns
456 (Supplementary Fig. 34) with uneven representation of each cell type.

457 We calculated the contribution of each cell type to the overall co-expression pattern observed.
458 We focused on the co-expression clusters, clusters C0 and C6, where the orthologs from the
459 extremophytes showed a constitutively-prepared expression comparable to the stress-induced or -
460 repressed expression of their counterparts in Arabidopsis (Fig. 6d). In ABA co-expression cluster
461 0 (4%-11% of the total clustered ABA-responsive orthologs in different cell types), where
462 extremophyte orthologs exhibited consistently higher expression both before and after ABA
463 exposure that those in the non-extremophytes, all cell types contributed evenly to this
464 constitutive expression pattern (Fig. 6d, upper left). Conversely, in cluster 6 (4%-16% of the total
465 clustered ABA-responsive orthologs in different cell types), characterized by the constitutively
466 low expression in the extremophytes, the predominant contributors were xylem cells (Fig. 6d,
467 upper right). In response to salt, the high constitutive expression of the extremophytes in cluster
468 0 (6%-14% of the total clustered NaCl-responsive orthologs in different cell types) was primarily
469 attributed to lateral root cap and atrichoblast cells, whereas the consistently low expression of the
470 extremophytes in cluster 6 (8%-16% of of the total clustered NaCl-responsive orthologs in
471 different cell types) was predominantly contributed by cortex and pericycle cells (Fig. 6d, bottom
472 panel). The co-expressed orthologs from lateral root cap and atrichoblast in cluster 0 under salt
473 were both predominantly enriched for RNA metabolism- and ribosome biogenesis-related
474 processes (Supplementary Fig. 35), which were also among the most responsive processes in
475 Arabidopsis outer cell layers (Fig. 5d). This underscores the tendency for certain cell types to
476 play a more substantial role in shaping the collective stress-prepared signal, as opposed to an
477 equal contribution from all cell types.

478 Moreover, we noticed that only a subset of, but not all, surface cell types, including LRC and
479 atrichoblasts, exhibited disproportionately higher contribution to the organ-level stress-prepared
480 state upon salt treatment. However, in response to ABA, more internal cell types, including
481 cortex and pericycle cells, showed stronger contributions to the stress-prepared signal than the
482 surface cell types (Fig. 6d). These results suggest that the contribution of individual cell types to
483 organ-level stress preparedness is likely shaped by a combination of their spatial position within

484 the tissue, the nature of the stress, and the cell type-specific capacities for stress perception and
485 signal transduction.

486 **Cell type assignments and sub-genome dominance at cell-type specificity in a polyploid**

487 To assess the generality and transferability of our comparative single-cell framework developed
488 for diploid species, we next applied our approach to the polyploid *Camelina sativa*. Unlike the
489 four diploid species analyzed above, *C. sativa* contains three closely related sub-genomes,
490 providing an opportunity to examine how cell-type identity and stress-responsive transcriptional
491 programs are distributed among sub-genomes and homeologs in a polyploid genomic context. *C.*
492 *sativa* is an allohexaploid in the Brassicaceae lineage I that has emerged as a promising oil crop
493 for sustainable biofuel production[42]. *C. sativa* can grow in diverse environments, but its overall
494 growth, seed oil yield, and oil quality are severely impacted by salt and other abiotic stressors[43–
495 45]. We generated a single cell atlas of 27,441 root cells from *C. sativa* under control, 5 μM ABA,
496 and 100 mM NaCl, as described for the other four species (Fig. 7). Using the cross-species cell-
497 type annotation strategy developed for this study, we were able to identify all ten major root cell
498 types in *C. sativa* (Fig. 7a), which exhibited a clear separation between inner and outer cell
499 layers as observed for other Brassicaceae species (Fig. 7b). The individual sub-genomes from *C.*
500 *sativa* did not significantly affect the integration between *C. sativa* and Arabidopsis or the cell-
501 type assignments (Supplementary Fig. 36), suggesting a relatively even representation of the
502 three sub-genomes. Our result is consistent with the undifferentiated sub-genome structure
503 previously reported for this species[46,47]. Moreover, we found that sub-genome G3 consistently
504 exhibited a higher global expression compared to the other sub-genomes across cell types and
505 conditions, except for columella cells where sub-genome G1 showed the highest mean
506 expression under control and ABA-treated conditions (Supplementary Fig. 37). This agrees with
507 prior work that has established that the sub-genome G3 in *C. sativa* exhibited partial expression
508 dominance over the other two sub-genomes[48].

509 Although sub-genome expression dominance has been extensively examined in many polyploid
510 species, it still remains unknown whether sub-genomes differentially contribute to stress
511 response. To determine this, we identified DEGs from each sub-genome in each cell type
512 (Supplementary Data 6). Despite the presence of a species-specific response, the impact of stress
513 on cell type transcriptomes and the functional processes represented by the DEGs in *C. sativa*
514 largely resembled the other four species (Figs. 7c-7d and Supplementary Fig. 38). The proportion
515 of DEGs did not significantly differ between sub-genomes (0.2%-2.4% and 1%-25% of
516 expressed genes across cell types for ABA and NaCl treatments, respectively), irrespective of
517 whether we considered all detected genes within individual sub-genomes or only complete
518 homeologs (defined as triplets)[48] (Supplementary Fig. 39). Consistently, transcriptomic
519 responsiveness measured as AUC scores did not differ significantly among the sub-genomes for
520 most cell types under ABA or NaCl treatments (Fig. 7e). These findings show that, in contrast to
521 the expression bias observed among sub-genomes across cell types under stress-neutral
522 conditions, stress responses did not further differentiate the sub-genomes in *C. sativa*, suggesting
523 a lack of large-scale sub-genome dominance in stress responses.

524 The recent polyploidization of the *C. sativa* genome[46] allowed us to ask what genomic changes
525 corresponded to divergence in gene expression across the three copies of each homeolog (18,565
526 complete triplets). First, to quantify coding-sequence divergence, we calculated the
527 nonsynonymous/synonymous rate ratio (ω) for each complete triplet. Second, to assess the
528 regulatory sequence divergences among homeologs, we considered divergence of expression at
529 three different parameters: basal level expression, cell-type preferential expression, and
530 expression changes in response to treatment. This analysis identified triplets exhibiting high
531 divergence (top 1% for each metric) among homeologs in 1) coding sequences (171 triplets), 2)
532 basal expression level (1,384 triplets), 3) expression specificity (115 triplets), and 4) stress
533 responses (140 triplets) (Fig. 7f; Supplementary Data 9). We found that these highly divergent
534 triplets were often shared across different cell types within each defined category, particularly for
535 these showing divergent basal expressions. Few triplets were divergent across multiple
536 categories with only 6 out of the 171 fast evolving triplets also showing high expression
537 divergence (overlap between group ① and groups ② and ⑤ in Fig. 7f). These results indicate
538 that coding-sequence and expression level divergences among triplets are largely uncoupled
539 from one another in *C. sativa*.

540 **DISCUSSION**

541 Comparing stress responses at the cellular level between extremophytes and non-extremophytes
542 holds the potential to identify innovations in stress resilience that may expedite the development
543 of climate-change ready crops. Here, we selected five closely related Brassicaceae species with
544 distinct responses to ABA and NaCl and generated their root transcriptomes at single cell
545 resolution. We developed a two-step approach for cross-species and cross-condition cell-type
546 annotations and confirmed the resulting annotation by *in situ* hybridization of selected markers
547 preferentially expressed in specific cell types, conditions, or species. In addition to serving as a
548 cell atlas resource for Brassicaceae roots under control and stress conditions, our method of cell-
549 type identity and response assessments under induced stressed conditions serves as a model for
550 other multi-species and treatment comparisons. Our approach was further applied on an
551 allohexaploid oilseed crop, *C. sativa* to demonstrate the versatile detection and transfer of cell-
552 type response features between diploid and polyploid genomes. We anticipate these data and the
553 methods presented in our study will provide new opportunities and perspectives for comparative
554 single cell transcriptomics to enrich our understanding of cell-type-specific evolution across
555 species. The searchable database we established to disseminate these data will facilitate their
556 broader use by the community. We envision that this resource will serve as a foundation for
557 downstream functional and experimental studies, especially in a phylogenetic context. It not only
558 facilitates targeted visualization and manipulation of cell types of interest, especially in species
559 where such information is currently limited, but also guides the selection of candidate genes for
560 testing cell type-specific stress responses across species.

561 To date, experimentally validated root cell-type markers in plants have been largely derived from
562 model species like Arabidopsis[49,50]. However, the transferability of these markers to other
563 species where root cell-type markers are unavailable has not been determined. We evaluated the
564 conservation of previously established Arabidopsis cell-type markers in the selected

565 Brassicaceae species and found that only 75 of the 241 Arabidopsis markers could serve as cell-
566 type markers across all species tested. The main basis for this lack of conservation was the low
567 expression levels among orthologs in related species. This suggested that moderate to high
568 expression of orthologs, in addition to preferential expression in the target cell type, should be an
569 important practical feature of widely used cell-type markers. We have newly identified over 300
570 cell-type preferentially expressed genes that are conserved across these Brassicaceae species
571 (Supplementary Data 3), similar to the resource available for grass species[39]. A previous study
572 comparing rice and Arabidopsis demonstrated that marker genes for the majority of root cell
573 types were poorly conserved[20]. This raises interesting questions about the relationship between
574 conservation of cell-type markers and phylogenetic distances among species. For example, at
575 which evolutionary distances do cell-type marker conservation erode and how do genes lose or
576 gain cell-type specificity? Advances in our understanding of cell type evolution will inform
577 innovations in synthetic biology, particularly cell-type specific engineering of traits or metabolite
578 production across species.

579 Identifying new cell populations across species enabled by scRNA-seq provides an
580 unprecedented opportunity to trace distinct cell states or origins of cell types and infer cell-type
581 evolution[51]. The presence of an extra cortex cell layer specific to *S. irio* and *S. parvula* root
582 anatomy was recapitulated with corresponding scRNA-seq data, which suggested the additional
583 cortex layer is transcriptionally specialized for flavonoid biosynthesis. Such evolutionary
584 innovations in cell layer compartmentalization may enable the spatial partitioning of growth and
585 stress-relevant metabolic functions. However, the exact link between this biased metabolic
586 partitioning and plant growth and stress responses remains unresolved. A second layer of cortex,
587 called middle cortex, can be also detected in Arabidopsis[52,53], but at a later developmental stage
588 of roots. Whether the inner cortex layer in *S. irio* and *S. parvula* is developmentally and
589 functionally similar to the Arabidopsis middle cortex, whether the flavonoid pathway is also
590 preferentially partitioned to middle cortex in Arabidopsis, or whether the early development of a
591 second cortex layer presents an evolutionary derived trait remains to be elucidated.

592 Beyond the comparisons under stress-neutral conditions, our dataset enables interrogation of
593 stress responses in each cell type across species. Our results demonstrate that environmental
594 stresses trigger distinct, cell-type- and stress-specific changes in transcriptome divergence across
595 species, with cell types exhibiting differential contributions to stress-responsive co-expression
596 networks. These findings highlight the distinct roles of cell types in driving stress response
597 divergence among species and the approaches developed here can therefore facilitate the
598 identification of cell types that play more significant roles in determining the divergence.

599 The cell-type responses of *C. sativa* highlighted transcriptional processes that are shared as well
600 as unique to root cell types in the polyploid oil crop. The emerging diversity of cell-type
601 responses shows the necessity of examining stress responses of crop species to assess the
602 transferability of response mechanisms from model plants. Although the sub-genome G3
603 demonstrated a greater contribution of highly expressed genes across the majority of cell types,
604 the three sub-genomes within this allohexaploid have not yet differentiated to the extent of
605 contributing to distinct stress responses in any single cell type. This lack of sub-genome stress

606 response dominance suggested that all three sub-genomes in *C. sativa* contribute comparably to
607 stress responses. Such distributed regulatory contributions may increase robustness by providing
608 functional redundancy among homeologs and buffering against divergence or loss of individual
609 gene copies. However, we do not infer that balanced sub-genome contributions directly enhance
610 stress tolerance, rather, these results indicate that *C. sativa* does not restrict stress-responsive
611 regulation to a single genomic origin. Our findings on response divergence rarely co-occurring
612 with sequence and expression divergence within individual homeolog groups suggest the
613 evolution of these divergences occurs largely independently. Together, these findings add
614 insights into the coordination among sub-genomes in polyploid species and the fate of
615 homeologs, and how we may generate agriculturally preferable neopolyploid crops[54].

616

617 **METHODS**

618 **Plant materials and growth conditions**

619 *Arabidopsis thaliana* (Arabidopsis, ecotype Col-0), *Eutrema salsugenium* (ecotype Shandong)
620 and *Schrenkiella parvula* (ecotype Lake Tuz) seeds were surface sterilized using a 30% (v/v)
621 bleach, 0.1% (v/v) Triton X-100 solution for 10 min and rinsed with sterile water 5 times.
622 *Sisymbrium irio* (original accession from Chris Pires, University of Missouri) seeds were
623 sterilized using 95% EtOH for 5 min, then 20% (v/v) bleach, 0.1% (v/v) Tween 20 solution for 5
624 min, and rinsed with sterilize water 4 times. *Camelina sativa* (ecotype DH55) seeds were surface
625 sterilized using 20% (v/v) bleach, 0.1% (v/v) Triton X-100 solution for 15 min and rinsed with
626 sterile water 4 times. After seed sterilization, seeds were incubated at 4°C for 2~3 days prior to
627 plating. Seeds were placed densely in 4~5 rows and incubated vertically on standard media (¼×
628 Murashige and Skoog (MS) nutrients, 1% sucrose, 0.05% MES, 1% agar, adjusted to pH 5.7
629 with 1 M KOH) covered with 100/47 µm sterilized mesh and incubated under continuous light at
630 22°C for 3 days (*Camelina sativa*) or 4~6 days (Arabidopsis, *Eutrema salsugenium*, *Schrenkiella*
631 *parvula* and *Sisymbrium irio*) until roots were ~2-3 cm in length. For ABA treatment, the mesh
632 with the seedlings was transferred to plates containing either 5 µM ABA or standard media (as
633 control) and incubated for 3 hrs. For NaCl treatment, the mesh with the seedlings was transferred
634 to plates containing either 100 mM NaCl or standard media (as control) and incubated for 24 hrs.

635 **Preparation of root samples for scRNA-seq**

636 After seedling growth on plates containing ABA, NaCl, or standard media, primary root tips,
637 which approximately corresponded to the top one third (0.7-1 cm) from the root tips, were
638 harvested from each species. The harvested root tips included root cap, meristematic, elongation,
639 and differentiation zones. Harvested root tips were placed into a 35 mm petri dish containing a
640 70 um strainer and 4 mL enzyme solution (1.25% [w/v] Cellulase ["ONOZUKA" R-10, Yakult],
641 0.1%[w/v] Pectolyase [P-3026, Sigma-Aldrich], 0.4 M Mannitol, 20 mM MES [pH 5.7], 20 mM
642 KCl, 10 mM $CaCl_2$, 0.1% [w/v] bovine serum albumin). The sample dish was rotated at 85 rpm
643 for 50 min at 25°C. The root tips were then transferred to a glass slide and gently pressed with
644 another glass cover using a pencil rubber to release protoplasts. The tissues were transferred back

645 to the enzyme solution and incubated for another 10 minutes. The protoplast solution was filtered
646 through a 70 µm filter, and then through a 40 µm filter twice using 50 mL conical tubes. The
647 filtered solution was centrifuged at 500g for 10 min at 22°C using a swinging bucket rotor and
648 the supernatant was discarded. Protoplasts were resuspended using 500 µL Solution A (0.4 M
649 Mannitol, 20 mM MES [pH 5.7], 20 mM KCl, 10 mM $CaCl_2$, 0.1% [w/v] bovine serum albumin)
650 and transferred to a 2 mL round bottom tube. After centrifugation at 200g for 6 min, the pelleted
651 protoplasts were resuspended with 30–50 µL Solution A to achieve the desired cell concentration
652 (~1,000 protoplasts/µL).

653 **scRNA-seq library construction and sequencing**

654 The protoplast suspension was loaded into Chromium microfluidic chips (Next GEM Single Cell
655 3' kit) and barcoded with a 10X Chromium Controller (10X Genomics). mRNA from the
656 barcoded cells were subsequently reverse-transcribed and sequencing libraries constructed with
657 reagents from a Chromium Single Cell 3' reagent kit v3.1 chemistry dual index (10X Genomics)
658 according to the manufacturer's instructions. The quality of cDNA and final library was assessed
659 with a Bioanalyzer (Agilent). Paired-end sequencing was performed with Illumina NovaSeq
660 6000 according to the manufacturer's instructions (Illumina).

661 **Bulk RNA-seq library preparation and sequencing**

662 Seedlings were grown under the same stress-neutral conditions as the scRNA-seq samples. RNA
663 was collected from the lower 1/3 of primary roots (which includes the entire developing root tip)
664 from each plant species using the RNeasy Plant Mini Kit (Qiagen). Library construction and
665 sequencing were performed by the University of Michigan Sequencing Core using the Illumina
666 TruSeq Kit followed by paired-end sequencing on Illumina NovaSeq 6000. Three replicates were
667 obtained per species.

668 **Orthologous gene sets**

669 We used two different sets of orthologous genes in our study: (1) comparative analyses among
670 Arabidopsis, *E. salsugenium*, *S. irio*, and *S. parvula* were performed using 15,198 unambiguous
671 1-to-1 ortholog groups between the four species. These 1-to-1 orthologues were identified by
672 comparing the primary protein-coding sequences and their syntenic positions using CLfinder[55]
673 with MMSeq2[56] as the aligner with default settings. (2) Species-specific analyses were
674 performed using all orthologs expressed in the given species. The orthologs between each non-
675 Arabidopsis species (including *E. salsugenium*, *S. irio*, *S. parvula*, and *C. sativa*) and
676 Arabidopsis were identified using the following steps: orthologous gene pairs of the protein-
677 coding genes between each non-Arabidopsis species and Arabidopsis were determined
678 reciprocally using MMseqs2[56] and were subsequently merged to generate a non-redundant list.
679 Only orthologous gene pairs that were considered as orthologs by both OrthoFinder (v2.5.4)[57]
680 with granularity -I set to 1.6 and CLfinder were kept. This filtered list was supplemented with
681 reciprocal best hits identified by blastp[58] to generate the final non-redundant orthologues
682 between each non-Arabidopsis species and Arabidopsis.

683 **Raw reads processing**

684 Raw single-cell RNA-seq FASTQ files were analyzed by Cell Ranger (v5.0.1, 10X Genomics) to
685 generate gene-by-cell count matrices. For each species, we retrieved the reference genome and
686 corresponding annotation as follows: Arabidopsis (TAIR10) from The Arabidopsis Information
687 Resource (TAIR; https://www.arabidopsis.org/), *C. sativa* (v2) from Brassicaceae Database
688 (BRAD; http://brassicadb.cn), *E. salsugineum* (v1.0) and *S. parvula* (v2.2) from Phytozome 12
689 (https://phytozome-next.jgi.doe.gov/) (Phytozome ID 574 and 173, respectively), reference
690 genome for *S. irio* (v1.0) from Brassicaceae Database (BRAD; http://brassicadb.cn) with the
691 updated gene model annotation[9] from the CoGe database (https://genomevolution.org/) with the
692 Genome ID 57216. To account for the low accuracy of 3' UTR annotation in non-Arabidopsis
693 species, we extended the 3' UTR of all genes in these species by 500bp, as reported before[59].
694 Reads from each species were aligned to the corresponding reference genome and counts were
695 quantified with the Cell Ranger 'count' command with default parameters.

696 **Comparison of gene expression from bulk and pseudo-bulk samples**

697 Raw 150-bp paired-end reads were quality checked with FastQC (v0.11.5;
698 https://www.bioinformatics.babraham.ac.uk/projects/fastqc) and aligned to the respective
699 genomes with HISAT2 (v2.2.1)[60]. Gene counts were calculated using featureCounts (v2.0.3)[61]
700 with the modified gene annotations. Pseudo-bulk gene counts from each scRNA-seq sample
701 were generated by aggregating Cell Ranger-derived counts for all filtered cells. We next
702 implemented edgeR (v3.36.0)[62] to identify genes that exhibited differential expression (FDR ≤
703 0.001, absolute $\log_2$FC ≥ 2) between the bulk and pseudo-bulk samples to be aware of any genes
704 that may deviate between how the data was generated. The list is provided in Supplementary
705 Data 10.

706 **Identification of doublets**

707 Doublets in each scRNA-seq dataset were detected with DoubletFinder (v3)[63] and subsequently
708 removed prior to any downstream analyses. Three input parameters are required for doublets
709 prediction: the number of expected real doublets (nExp), the number of artificial doublets (pN)
710 and the neighborhood size (pK). We selected a stringent nExp of 7.5% based on the estimated
711 doublet rate in 10X Genomics Chromium Single Cell 3' Reagent Kit User Guide, which was
712 subsequently adjusted according to the proportion of homotypic doublets and doublet ratio. For
713 pN, the default value 25% was used as DoubletFinder performance is largely invariant of pN[63].
714 The optimal pK value was determined by sequentially loading pre-processed Seurat [22] data into
715 the "paramSweep_v3 (PCs = 1:15)", "summarizeSweep", and "find.pK" functions. For each
716 sample from each species, the single easily discernible maximum pK value was selected as the
717 optimal pK parameter. Only cells which were flagged as singlets were used for downstream
718 analyses.

719 **Data integration, clustering**

720 Data integration and clustering were performed using R package Seurat (v4.0.5)[22]. Prior to data
721 integration, we further removed cells in which the numbers of expressed genes were less than
722 400 and more than 10,000 or < 5% mitochondrial reads, and genes expressed in < 3 cells.

723 For integration of datasets from each species, we first log-normalized each dataset using the
724 *NormalizeData* function and identified the top 2,000 highly variable genes (HVGs) using a
725 variance-stabilizing transformation in *FindVariableFeatures*. We then used
726 *SelectIntegrationFeatures* to determine 2,000 integration features, which were subsequently used
727 in *FindIntegrationAnchors* to decide integration anchors between datasets. Datasets from each
728 species were eventually integrated using *IntegrateData* with 30 principal components (PCs). The
729 integrated dataset for each species was scaled and reduced dimensionally with Principal
730 Component Analysis (PCA) followed by Uniform Manifold Approximation and Projection
731 (UMAP) using 50 PCs. *FindNeighbors* and *FindClusters* were used to determine the nearest
732 neighbors and clusters, respectively.

733 For global integration of datasets from all species, feature names in each of the individual
734 stringently filtered datasets were replaced by their corresponding ortholog group ID. Each
735 dataset was normalized using the *SCTransform* function, and then *SelectIntegrationFeatures* was
736 used to identify 3,000 integration features. Integration features were combined using
737 *PrepSCTIntegration* and integration anchors were selected using *FindIntegrationAnchors*.
738 Datasets from all species were then integrated at once using the Seurat "SCT" approach in
739 *IntegrateData* function with setting Arabidopsis control datasets as the reference. Finally,
740 dimensional reduction of the integrated dataset was performed using the UMAP algorithm with
741 the top 100 PCs identified by PCA.

742 Dimensionality-reduction methods were used for distinct analytical purposes in our study.
743 UMAP was employed for visualization, integration, and cell-type annotation of single-cell data,
744 where preservation of local neighborhood structure is advantageous. In contrast, PCA was
745 applied to pseudo-bulk cell-type transcriptomes for cross-species and cross-condition
746 comparisons, as PCA preserves global variance structure and enables direct interpretation of
747 transcriptomic similarity and divergence. Because UMAP embeddings are nonlinear, stochastic,
748 and sensitive to parameter choices, and they are not used for quantitative distance- or variance-
749 based comparisons.

750 **Cell type assignment**

751 **Step 1. Establishment of a reference atlas using Arabidopsis control samples**

752 *Correlation-based annotation* (Step1 - M1.1 in Supplementary Fig. 4): We obtained bulk RNA-
753 seq data[64] and ATH1 microarray data[65] for 14 Arabidopsis root cell types isolated with
754 fluorescence activated cell sorting (FACS) as fragments per kilobase of transcript per million
755 fragments mapped (FPKM) and normalized probe values, respectively. We then computed the
756 expression variance for all genes within each dataset to identify highly variable genes (HVGs)
757 and categorize them into percentile bins. We hypothesized that the transcriptome signature of an
758 individual cell should remain most similar to that of the cell type it originates from regardless of

759 the HVGs used. Therefore, we calculated the Pearson correlation coefficient between each cell
760 from Arabidopsis control samples and the whole-transcriptome reference expression profiles for
761 cell types using HVGs from top 10% to 100% percentile bins. At each percentile bin, each cell
762 was annotated as the cell type with the highest correlation value.

763 *Marker-based annotation* (Step1 - M1.2 in Supplementary Fig. 4): We retrieved the normalized
764 counts and uniform manifold approximation and projection (UMAP) embeddings in Seurat for
765 all cells from Arabidopsis control samples, which were used to calculate the enrichment scores
766 of all genes for each cell using a cluster-independent approach in SEMITONES[66]. Similarity
767 between cells was estimated by a radial basis function kernel over the multidimensional UMAP
768 space, and the enrichment scores for all genes were computed using a simple linear regression
769 framework. We subsequently extracted the enrichment scores for 454 (after removing stele
770 markers) known Arabidopsis root cell type marker genes (Supplementary Data 2). This set was
771 compiled through an extensive literature review[15–21]. To identify statistically significantly
772 enriched markers for individual cells, an empirical distribution of enrichment scores for each cell
773 was obtained by permutating expression data 100 times. A gene is considered significantly
774 enriched in the given cell if its enrichment score is more than n standard deviations (nsds) away
775 from the mean of the empirical distribution. We determined the significantly enriched markers at
776 nsds of 5, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50 for each cell, and annotated the cell according to
777 the marker with the highest enrichment score at each nsds.

778 *Integration-based annotation* (Step1 - M1.3 in Supplementary Fig. 4): We used a previous
779 Arabidopsis root single-cell study[21] with cell type annotations for individual cells.
780 *FindTransferAnchors* from Seurat was used to find transfer anchor cells using the published data
781 as the reference and cells from Arabidopsis control samples in our dataset as the query.
782 *TransferData* was subsequently employed to predict the identity of each cell in the query dataset.
783 These steps were run with different thresholds for the key parameters: ndim at 30, 50 to 500 with
784 an increment interval of 50; kscore at 10, 30, 50; kweight at 10, 20, 30, 50, 80, 100. Cell type
785 prediction for each cell was extracted at every parameter combination.

786 *Combination of annotation methods* (Step1 in Supplementary Fig. 4): We first determined the
787 consensus annotation in each of the three annotation approaches for each cell by identifying the
788 most recurring (typically over 50%) annotation from all the runs. The final annotation for each
789 cell in Arabidopsis control samples was assigned by combining the information from the three
790 annotation approaches. If a cell was annotated to the same cell type from at least two of the three
791 approaches, then it was assigned that cell type. In rare cases where a cell cannot be annotated
792 using a consensus, the cell type was assigned according to the marker-based annotation.

793 **Step 2. Annotation transfer from Arabidopsis to other species**

794 The integration of cells from different treatments and different species in our dataset suggested
795 there were no treatment- or species-specific cell populations. We next employed three different
796 reference-based cell type annotation algorithms, including integration-based prediction (Step2 -
797 M2.1 in Supplementary Fig. 4) in Seurat[22], clustering-based prediction (Step2 - M2.2 in
798 Supplementary Fig. 4) in Symphony[23], and the machine learning-based prediction (Step2 - M2.3

799 in Supplementary Fig. 4) in scArches[24], to transfer cell type annotation from Arabidopsis control
800 samples to treatment samples and samples from other four species. To avoid probabilistic
801 annotations in these methods, we varied the key parameters in each method: Seurat was run with
802 ndim at 30, 50 to 500 with an increment interval of 50, kscore at 10, 30, 50, kweight at 10, 20,
803 30, 50, 80, 100; Symphony was run with hvg_num from 2000 to 14000 with an interval of 2000,
804 pc_num from 20 to 100 with an interval of 20, and k_num at 5, 10, 20, 30, 50, 80, 100; and
805 scArches was run with hvg_num from 2000 to 14000 with an interval of 2000. Similar to the
806 approach we used to annotate cells in the Arabidopsis reference, we next identified the most
807 recurring cell type annotation for a cell as the consensus annotation for that cell in each of the
808 three methods. The annotation for a cell with the highest votes from all three methods was
809 assigned as the final annotation for that cell. In rare cases where a cell was annotated differently
810 from the three methods, we compared the average probability given by each method and
811 assigned the cell to the prediction with the highest probability.

812 **Cell-type marker gene identification**

813 To identify cell type marker genes for each species, we used the Seurat *FindMarkers* function on
814 scaled expression using the default Wilcoxon test. Significant markers were selected based on
815 the following criteria: adjust p-value ≤ 0.01, average log fold change ≥ 1, cell type specificity $\geq$
816 0.9 and pct.diff > 0.75, where pct.dff is the ratio of the difference in the percentage of cells
817 expressing the gene between the target cell type and the rest of the cells to the percentage of cells
818 expressing the gene in the target cell type. To examine the expression of the cell-type markers,
819 the cell-type pseudo-bulk expression was computed by normalizing the raw gene counts in each
820 cell by the total counts of that cell, and these normalized counts were then averaged across all
821 cells belonging to the same cell type. This cell-level normalization ensures that pseudo-bulk
822 profiles reflect the average transcriptional state of individual cells rather than being dominated
823 by differences in cell numbers or library size.

824 **Cell type specificity indexes**

825 Cell type specificity index (τ) is based on the Tau metric of tissue specificity[67] as following:
826 $\sum_{i=1}^{N}(1 - x_i)/(N - 1)$ where N is the number of cell types and x_i is the expression level in cell
827 type normalized by the maximal expression of the given gene across all cell types. For this
828 analysis, the normalized expression as described for generating cell-type marker expression was
829 used.

830 **Identification of orthologues with divergent expression**

831 To identify divergently expressed orthologues (Fig. 3b), we compared gene expression of the
832 15,198 1-to-1 orthologues along the defined cell types across Arabidopsis, *E. salsugenium*, *S.*
833 *irio*, and *S. parvula*. Within each species, we determined the genes that were differentially
834 expressed between any pairs of cell types with Seurat *FindMarkers* function. We then compiled
835 these genes from all four species and used the normalized cell-type pseudo-bulk expression (the
836 cell-type pseudo-bulk expression was computed using the same approach as that used for cell
837 type specificity and then z-score normalized across cell types) of these genes as the input to
838 cluster gene expression along the ten cell types using an unsupervised soft clustering method

839 implemented in the R package Mfuzz (v.2.54.0)[68], which assigns probabilistic membership of each
840 ortholog to expression-pattern clusters rather than enforcing hard classifications. The number of
841 clusters and the fuzzification parameter required by Mfuzz were estimated by the Dmin and
842 mestimate functions. Subsequently, we inferred within a phylogenetic framework, as described
843 by Cardoso-Moreira *et al*.[69], the probability that there were changes in the expressions of the 1-
844 to-1 orthologues along the defined cell types, that is, that genes changed their cluster assignment
845 in specific branches, using 0.001 as the probability cut-off. This probability-based threshold
846 followed the original framework[69] and was selected to prioritize robust expression shifts over
847 marginal or noisy changes. In this framework, the ancestral cluster probabilities along the
848 phylogenetic tree were calculated as the weighted averages from the child-nodes. The weights
849 are given by the inverse branch lengths, which were retrieved from a previous study [13] using
850 WebPlotDigitizer (https://automeris.io/WebPlotDigitizer), so that closer child-nodes have more
851 weight. Since this framework relies only on quantitative expression measurements across
852 homologous biological contexts and a resolved species phylogeny, it is agnostic to organismal
853 lineage and can be applied to any systems that meet these criteria.

854 To construct the phylogeny for aquaporin subfamily NIPs and Tau class glutathione S-
855 transferase GSTUs across species, we first identified members from these families in
856 Arabidopsis from previous studies[70,71], and extracted their orthologs from *E. salsugenium*, *S. irio*,
857 and *S. parvula* using the ortholog network generated from CLfinder[55]. The protein sequences of
858 genes in each subfamily were aligned using MAFFT[72], the results of which were subsequently
859 used for phylogenetic tree construction with FastTree[73] with default settings. The inferred
860 phylogenies were annotated with ggtree[74] in R.

861 **Prioritization of cell types responsive to ABA and NaCl**

862 To estimate the degree of transcriptomic change for each cell type under treatments (Fig. 5b), we
863 used Augur[38] to rank cell types based on their degree of response to biological perturbations (i.e.
864 ABA or NaCl treatments in our current study). Augur uses a machine-learning frame to quantify
865 the separability of perturbed and unperturbed cells as AUC scores, based on the assumption that
866 cell types more responsive to a perturbation should be more separable than less affected cell
867 types within high-dimensional single-cell measurements. Augur analysis was performed for
868 ABA- and NaCl-treated samples separately compared to the control within each of the five
869 species. The significance of cell type prioritization for each cell type was determined by
870 comparing the AUC scores from all subsamples of the specific cell type of interest to all rest of
871 cell types for each treatment using one-tailed Wilcoxon rank-sum tests[75].

872 To compare cell-type transcriptomic responses across Arabidopsis, *E. salsugenium*, *S. irio*, and
873 *S. parvula*, we modified the Augur method to estimate transcriptome divergence of each cell type
874 between different species under each condition. Specifically, instead of using cells from different
875 conditions, we fed cells from different species under the same condition into Augur. This process
876 was repeated for all six pairwise combinations from the four species under each condition. For
877 each pair of species, the AUC scores of each cell type from ABA- and NaCl-treated samples
878 were compared to these from the control using Student's t-Test.

879 **Pseudo-bulk differential expression analysis**

880 Recent benchmark studies to assess methods to determine differentially expressed genes (DEGs)
881 for scRNA-seq data have shown that pseudo-bulk methods, which aggregate cell-level counts
882 into sample-level "pseudo-bulk" counts, outperform single-cell DE methods and achieve highest
883 fidelity to the experimental ground truth[76,77]. Therefore, we employed a pseudo-bulk approach to
884 detect genes that responded to the treatments for each cell type in each species. Pseudo-bulk
885 transcriptomes were aggregated for each of the cell types by taking the sum of raw cell-level
886 counts. Generalized linear model likelihood ratio test from edgeR (v3.36.0)[78] was performed
887 between the treatment and the control. For each comparison, cell types with less than 30
888 annotated cells under either treatment or control conditions were excluded from the differential
889 expression analysis. A gene was considered differentially expressed in a given cell population if
890 the false discovery-rate adjusted *p*-value was <= 0.01. Additionally, to calculate the
891 transcriptomic distance between samples, the raw counts in each cell type for the 1-to-1
892 orthologs were extracted from each species and each condition, and then were normalized in
893 DESeq2[79] as variance stabilized expression. The variance stabilized expression was then used for
894 PCA-based transcriptomic distance measurements.

895 **Identification of differential regulation of orthologs between species**

896 We extracted DEGs from each species under each treatment that were included in the 1-to-1
897 orthologs and compared their response to the treatment across species to determine differentially
898 regulated orthologs. We considered the stress response of the 1-to-1 orthologs as being
899 upregulated and downregulated if $\log_2FC > 0$ and $\log_2FC < 0$, respectively, regardless of their
900 response magnitude. To further distinguish observed differential regulation patterns between the
901 orthologs which occurred above chance due to random sampling (Fig. 6a), we started with a
902 fixed order of all 1-to-1 orthologs and then permuted the order of DEGs for 10,000 times
903 independently in each species for each cell type and generated an expected occurrence
904 distribution for each pattern when assuming independence between the species. The deviation of
905 the observed occurrence for each pattern was computed as the number of standard deviations
906 away from the mean of the corresponding expected distribution. The *p*-value of the observed
907 occurrence for each pattern was calculated by fitting the observed occurrences into the
908 corresponding bootstrapped distributions using R function *pnorm* and subsequently corrected for
909 multiple testing using the Benjamini-Hochberg procedure in R function *p.adjust*. To additionally
910 test the impact of gene duplication on the differential regulation of orthologs, we identified
911 ortholog groups with more than one copy in only one species where identification of correct
912 corresponding orthologs across species is feasible. The same approach was applied to these
913 duplicated ortholog groups to determine the deviation between the occurrences and the
914 expectations.

915 To compare the differential regulation among orthologs to their ABF binding events, we
916 retrieved DAP-seq data of four ABFs in Arabidopsis, *E. salsugenium*, *S. irio*, and *S. parvula*
917 from a previous study[9]. For each of the six possible species pairs, we extracted ortholog pairs
918 where at least one gene responded to the treatments and at least one gene in the pair contained a
919 minimum of one ABF binding event in its promoter region. Two kilobases upstream of the

920 transcription start sites were used as the searched promoter sequence space for ABF binding. We
921 then categorized these ortholog pairs into the following two groups: 1) orthologs with conserved
922 ABF binding where both genes contained at least one ABF binding event in their promoters; and
923 2) orthologs with divergent ABF binding where only one gene contained at least one ABF
924 binding event in its promoter. From these groups, we identified orthologs with only one gene
925 responding to the treatments in each cell type, which we categorized as divergently regulated
926 orthologs. The significance of differences in the percentage of divergently regulated orthologs
927 between the two defined groups were determined with *t.test* function in R.

928 **Determination of cell type contribution to co-expression patterns**

929 For each cell type under each treatment, orthologs that were considered differentially expressed
930 in at least one of the four species were extracted, and subsequently combined for aggregated
931 clustering. The normalized pseudo-bulk cell type expression levels of these orthologs were
932 subjected to fuzzy k-means clustering implemented in aerie[80]. Ortholog pairs in each of the
933 resulting clusters were further filtered based on: (1) the membership of a given ortholog pair was
934 no less than 0.5; (2) the Pearson correlation between the expression of a given ortholog pair and
935 the mean expression of the cluster it belongs to was no less than 0.8 with a *p*-value smaller than
936 0.01; and (3) the cluster needed to have a minimum of 50 members. To test the representation of
937 each co-expression pattern among cell types under each treatment, we compared the proportion
938 of orthologous DEGs derived from different cell types in each co-expression cluster to that
939 combined from all clusters using R function *phyper* with lower.tail = FALSE. The resulting *p*-
940 values were corrected for multiple testing using the Benjamini-Hochberg procedure in R function
941 *p.adjust*.

942 **Identification of extra cortex population**

943 To determine potential topological differences in cortex between species, we performed non-
944 linear dimensional reduction for cortex cells from each species under control condition using
945 Seurat[22] to generate species-specific UMAP representations. The UMAP representation of cortex
946 cells in *S. irio* was selected as the reference because of the presence of three distinct
947 subpopulations, which were subsequently annotated as left, middle, and right subpopulations
948 based on Monocle[81,82] trajectories. Genes that are differentially expressed in the three
949 subpopulations were identified using Seurat *FindMarkers* function with the default Wilcoxon
950 test.

951 Cortex cells from Arabidopsis, *E. salsugenium*, and *S. parvula* were mapped to the selected
952 reference, and assigned to reference left, middle, and right subpopulation using *MapQuery* in
953 Seurat[22]. Under- and over-representation of each subpopulation in these three species compared
954 to *S. irio* were determined by hypergeometric test using R function *phyper* with lower.tail =
955 TRUE and lower.tail = FALSE, respectively. The resulting *p*-values were corrected for multiple
956 testing using the Benjamini-Hochberg procedure in R function *p.adjust*.

957 To examine the expression of flavonoid biosynthesis genes in other species with more than one
958 layer of cortex, we've retrieved published scRNA-seq data for rice[33] and Medicago[34]. The
959 annotated cortex cells were extracted from these dataset and re-analyzed using Seurat[22]. The

960 orthologs of flavonoid biosynthesis genes from these two species were identified by comparing
961 their proteomes to the Arabidopsis proteome using OrthoFinder (v2.5.4)[57]. The expression of the
962 identified flavonoid biosynthesis orthologs were then examined in the extracted cortex
963 population for each species.

964 **Comparison of basal expression and stress responses among sub-genomes in *C. sativa***

965 To compare the contribution of sub-genomes to the data integration between Arabidopsis and *C.*
966 *sativa*, we used the latest refined sub-genome structure for *C. sativa* as well as the refined
967 syntenic orthologs between these two species[48]. The integration was performed separately with
968 all sub-genomes together as well as with each sub-genome separately (Supplementary Fig. 36)
969 using the default settings in IntegrateData from Seurat. For comparison of basal expression and
970 stress responses, we only considered triplets with three complete homeologs. The distribution of
971 DEGs among the three sub-genomes was tested using R function phyper with lower.tail =
972 FALSE. The resulting *p*-values were corrected for multiple testing using the Benjamini-
973 Hochberg procedure in R function p.adjust.

974 **Identification of fast diverging homeologous groups in *C. sativa***

975 We identified fast diverging homeologs at the following levels: 1) coding sequences, 2) basal
976 expression level, 3) expression specificity, and 4) stress responses. Divergence at the coding
977 level among homeologs was evaluated using the ω value (dN/dS), which was calculated for each
978 complete triplet using CODEML[83]. Any homeologous groups where either dN or dS was
979 estimated to be zero were removed from downstream analyses. Divergence in expression and
980 expression specificity were computed as the coefficient of variation of variance stabilized
981 expression derived from DESeq2[79] and Tau indexes of the homeologs within each homeologous
982 group. Similarly, we used the standard deviation of the fold changes of the homeologs as the
983 estimation for the divergence in stress responses for each triplet. For each of these categories,
984 only homeologous groups that fall in the top one percentile of the metrics, regardless of cell
985 types, were considered as the fast diverging homeologous groups.

986 **Gene ontology analysis**

987 Enrichment analysis of biological pathways on gene sets of interest was performed using
988 BiNGO[84]. Gene ontology (GO) annotations for Arabidopsis were retrieved from the GO
989 consortium (http://geneontology.org/) on December 12th, 2021. For non-Arabidopsis species, this
990 analysis was conducted either by replacing gene IDs with their Arabidopsis ortholog IDs (for *E.*
991 *salsugenium*, *S. irio*, and *S. parvula*) or annotating genes with GO annotations transferred from
992 Arabidopsis (for *C. sativa*). Redundant GO terms with > 50% overlap with similar terms were
993 further clustered using Markov clustering implemented via GOMCL[85]
994 (https://github.com/Guannan-Wang/GOMCL).

995 **Whole-mount *in situ* hybridization**

Seeds were stratified and then germinated on sterile 1x MS standard media (1X Murashige and Skooge nutrients (Caisson MSP01), MES hydrate 0.5g/L, 1% sucrose and 0.7% agar (Daishin CAS# 9002-18-0) at pH 5.7). The plates were incubated vertically in growth chambers (Percival Scientific CU-36L4) at 22°C with a 16/8 hour light/dark cycles and grown until the roots were ~2-3 cm in length (5 days for Arabidopsis, 4 days for *S. irio*, 4-5 days for *S. parvula*, and 6-7 days for *E. salsugineum*).

Probes for selected candidates were designed and synthesized by Molecular Instruments (https://www.molecularinstruments.com). Sample fixation, dehydration, permeabilization, protease digestion, probe hybridization and amplification were performed as described in Oliva *et al*.[31] with the following modifications: after hybridization and washes, roots with were left in ClearSee[86] overnight and then stained with 0.22% calcofluor solution (Flourescent Brightener 28, Santa Cruz Biotechnology sc-218504) to demarcate cell layers. Samples were mounted on glass slides and imaged with SP8 Leica Confocal Microscope, using a 20X oil immersion objective. Calcofluor white and Alexa 647-coupled amplifiers were excited at 405 nm and 640 nm, and collected at 425-475 nm and 650-675 nm, respectively. Leica LAS X Navigator was used to collect and merge tiled images wherever the differentiated zone was included. Images were processed using Fiji[87].

DATA AVAILABILITY

All data are available in the main text, the Supplemental Information, referenced databases and public databases. All raw sequence data generated in the current study have been deposited at NCBI under BioProject PRJNA1113801. All processed data can be found at NCBI GEO database under accession code GEO: GSE268881. A web browser to interactively explore the data from the current study is available at: https://plantbiology.shinyapps.io/brassicaceaeatlas.

CODE AVAILABILITY

All software used in this study is publicly available as described in the Methods. Detailed parameters used for analyzing each type of sequencing data have been described in the Methods. All original code used throughout the data analysis is publicly available at the following GitHub repository: https://github.com/Guannan-Wang/BrassicaceaeAtlas.

REFERENCES

1. Dinneny, J. R. *et al.* Cell identity mediates the response of Arabidopsis roots to abiotic stress. *Science* **320**, 942–945 (2008).

2. Iyer-Pascuzzi, A. S. *et al.* Cell identity regulators link development and stress responses in the Arabidopsis root. *Dev. Cell* **21**, 770–782 (2011).

3. Kawa, D. & Brady, S. M. Root cell types as an interface for biotic interactions. *Trends Plant Sci.* **27**, 1173–1186 (2022).

4. Zhang, H., Zhao, Y. & Zhu, J.-K. Thriving under Stress: How Plants Balance Growth and the Stress Response. *Dev. Cell* **55**, 529–543 (2020).

1033 5. Waadt, R. *et al.* Plant hormone regulation of abiotic stress responses. *Nat. Rev. Mol. Cell*
1034 *Biol.* **23**, 680–694 (2022).

1035 6. Xu, S., Wang, J., Guo, Z., He, Z. & Shi, S. Genomic Convergence in the Adaptation to
1036 Extreme Environments. *Plant Commun* **1**, 100117 (2020).

1037 7. Monihan, S. M. *et al.* Duplication and functional divergence of a calcium sensor in the
1038 Brassicaceae. *J. Exp. Bot.* **71**, 2782–2795 (2020).

1039 8. Kazachkova, Y. *et al.* Halophytism: What Have We Learnt From Arabidopsis thaliana
1040 Relative Model Systems? *Plant Physiol.* **178**, 972–988 (2018).

1041 9. Sun, Y. *et al.* Divergence in the ABA gene regulatory network underlies differential growth
1042 control. *Nat Plants* **8**, 549–560 (2022).

1043 10. Tran, K.-N. *et al.* Multiple paths lead to salt tolerance - pre-adaptation vs dynamic
1044 responses from two closely related extremophytes. *bioRxiv* 2021.10.23.465591 (2022)
1045 doi:10.1101/2021.10.23.465591.

1046 11. Geng, Y. *et al.* A spatio-temporal understanding of growth regulation during the salt stress
1047 response in Arabidopsis. *Plant Cell* **25**, 2132–2154 (2013).

1048 12. Tran, K.-N. *et al.* Balancing growth amidst salt stress - lifestyle perspectives from the
1049 extremophyte model Schrenkiella parvula. *Plant J.* **116**, 921–941 (2023).

1050 13. Huang, C.-H. *et al.* Resolution of Brassicaceae Phylogeny Using Nuclear Genes Uncovers
1051 Nested Radiations and Supports Convergent Morphological Evolution. *Mol. Biol. Evol.* **33**,
1052 394–412 (2016).

1053 14. Oh, D.-H. *et al.* Genome structures and transcriptomes signify niche adaptation for the
1054 multiple-ion-tolerant extremophyte Schrenkiella parvula. *Plant Physiol.* **164**, 2123–2138
1055 (2014).

1056 15. Ryu, K. H., Huang, L., Kang, H. M. & Schiefelbein, J. Single-Cell RNA Sequencing
1057 Resolves Molecular Relationships Among Individual Plant Cells. *Plant Physiol.* **179**, 1444–
1058 1456 (2019).

1059 16. Jean-Baptiste, K. *et al.* Dynamics of Gene Expression in Single Root Cells of Arabidopsis
1060 thaliana. *Plant Cell* **31**, 993–1011 (2019).

1061 17. Denyer, T. *et al.* Spatiotemporal Developmental Trajectories in the Arabidopsis Root
1062 Revealed Using High-Throughput Single-Cell RNA Sequencing. *Dev. Cell* **48**, 840–852.e5
1063 (2019).

1064 18. Wendrich, J. R. *et al.* Vascular transcription factors guide plant epidermal responses to
1065 limiting phosphate conditions. *Science* **370**, (2020).

1066 19. Coate, J. E., Farmer, A. D., Schiefelbein, J. W. & Doyle, J. J. Expression Partitioning of
1067 Duplicate Genes at Single Cell Resolution in Arabidopsis Roots. *Front. Genet.* **11**, 596150
1068 (2020).

1069 20. Liu, Q. *et al.* Transcriptional landscape of rice roots at the single-cell resolution. *Mol. Plant*
1070 **14**, 384–394 (2021).

1071 21. Shahan, R. *et al.* A single-cell Arabidopsis root atlas reveals developmental trajectories in
1072 wild-type and cell identity mutants. *Dev. Cell* **57**, 543–560.e9 (2022).

1073 22. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21
1074 (2019).

1075 23. Kang, J. B. *et al.* Efficient and precise single-cell reference atlas mapping with Symphony.
1076 *Nat. Commun.* **12**, 5890 (2021).

1077 24. Lotfollahi, M. *et al.* Mapping single-cell data to reference atlases by transfer learning. *Nat.*
1078 *Biotechnol.* **40**, 121–130 (2022).

1079 25. Choi, H. M. T. *et al.* Third-generation in situ hybridization chain reaction: multiplexed,
1080 quantitative, sensitive, versatile, robust. *Development* **145**, (2018).

1081 26. Takano, J. *et al.* The Arabidopsis major intrinsic protein NIP5;1 is essential for efficient
1082 boron uptake and plant development under boron limitation. *Plant Cell* **18**, 1498–1509
1083 (2006).

1084 27. Wang, G. *et al.* Cross species multi-omics reveals cell wall sequestration and elevated
1085 global transcript abundance as mechanisms of boron tolerance in plants. *New Phytol.* **230**,
1086 1985–2000 (2021).

1087 28. Kumar, S. & Trivedi, P. K. Glutathione S-Transferases: Role in Combating Abiotic Stresses
1088 Including Arsenic Detoxification in Plants. *Front. Plant Sci.* **9**, 751 (2018).

1089 29. Sappl, P. G. *et al.* The Arabidopsis glutathione transferase gene family displays complex
1090 stress regulation and co-silencing multiple genes results in altered metabolic sensitivity to
1091 oxidative stress. *Plant J.* **58**, 53–68 (2009).

1092 30. Terrón-Camero, L. C. *et al.* Gene network downstream plant stress response modulated by
1093 peroxisomal H2O2. *Front. Plant Sci.* **13**, 930721 (2022).

1094 31. Oliva, M. *et al.* An environmentally responsive transcriptional state modulates cell identities
1095 during root development. *bioRxiv* 2022.03.04.483008 (2022)
1096 doi:10.1101/2022.03.04.483008.

1097 32. Daryanavard, H., Postiglione, A. E., Mühlemann, J. K. & Muday, G. K. Flavonols modulate
1098 plant development, signaling, and stress responses. *Curr. Opin. Plant Biol.* **72**, 102350
1099 (2023).

1100 33. Zhu, M. *et al.* Single-cell transcriptomics reveal how root tissues adapt to soil stress. *Nature*
1101 (2025) doi:10.1038/s41586-025-08941-z.

1102 34. Cervantes-Pérez, S. A. *et al.* Cell-specific pathways recruited for symbiotic nodulation in
1103 the Medicago truncatula legume. *Mol. Plant* **15**, 1868–1888 (2022).

1104 35. Shomali, A. *et al.* Diverse physiological roles of flavonoids in plant environmental stress
1105 responses and tolerance. *Plants* **11**, 3158 (2022).

1106 36. Wu, J. *et al.* Advances in the study of the function and mechanism of the action of
1107 flavonoids in plants under environmental stresses. *Planta* **257**, 108 (2023).

1108 37. Farmer, A., Thibivilliers, S., Ryu, K. H., Schiefelbein, J. & Libault, M. Single-nucleus RNA
1109 and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in
1110 Arabidopsis roots at the single-cell level. *Mol. Plant* **14**, 372–383 (2021).

1111 38. Skinnider, M. A. *et al.* Cell type prioritization in single-cell data. *Nat. Biotechnol.* **39**, 30–34
1112 (2021).

1113 39. Guillotin, B. *et al.* A pan-grass transcriptome reveals patterns of cellular divergence in
1114 crops. *Nature* **617**, 785–791 (2023).

1115 40. Gong, Q., Li, P., Ma, S., Indu Rupassara, S. & Bohnert, H. J. Salinity stress adaptation
1116 competence in the extremophile Thellungiella halophila in comparison with its relative
1117 Arabidopsis thaliana. *Plant J.* **44**, 826–839 (2005).

1118 41. Lee, Y. P. *et al.* Salt stress responses in a geographically diverse collection of
1119 Eutrema/Thellungiella spp. accessions. *Funct. Plant Biol.* **43**, 590–606 (2016).

1120 42. Moser, B. R. Biodiesel from alternative oilseed feedstocks: camelina and field pennycress.
1121 *Biofuels* **3**, 193–209 (2012).

1122 43. Steppuhn, H., Falk, K. C. & Zhou, R. Emergence, height, grain yield and oil content of
1123 camelina and canola grown in saline media. *Can. J. Soil Sci.* **90**, 151–164 (2010).

1124 44. Nadakuduti, S. S. *et al.* Heat stress during seed development leads to impaired physiological
1125 function and plasticity in seed oil accumulation in Camelina sativa. *Front. Plant Sci.* **14**,
1126 1284573 (2023).

1127 45. Ahmed, Z. *et al.* Differential physio-biochemical and yield responses of Camelina sativa L.
1128 under varying irrigation water regimes in semi-arid climatic conditions. *PLoS One* **15**,
1129 e0242441 (2020).

1130 46. Kagale, S. *et al.* The emerging biofuel crop Camelina sativa retains a highly
1131 undifferentiated hexaploid genome structure. *Nat. Commun.* **5**, 3706 (2014).

1132 47. Blume, R. Y., Kalendar, R., Guo, L., Cahoon, E. B. & Blume, Y. B. Overcoming genetic
1133 paucity of Camelina sativa: possibilities for interspecific hybridization conditioned by the
1134 genus evolution pathway. *Front. Plant Sci.* **14**, 1259431 (2023).

1135 48. Chaudhary, R. *et al.* Assessing Diversity in the Camelina Genus Provides Insights into the
1136 Genome Structure of Camelina sativa. *G3* **10**, 1297–1308 (2020).

1137 49. Jin, J. *et al.* PCMDB: a curated and comprehensive resource of plant cell markers. *Nucleic*
1138 *Acids Res.* **50**, D1448–D1455 (2022).

1139 50. Grones, C. *et al.* Best practices for the execution, analysis, and data storage of plant single-
1140 cell/nucleus transcriptomics. *Plant Cell* **36**, 812–828 (2024).

1141 51. Wang, J. *et al.* Tracing cell-type evolution by cross-species comparison of cell atlases. *Cell*
1142 *Rep.* **34**, 108803 (2021).

1143 52. Baum, S. F., Dubrovsky, J. G. & Rost, T. L. Apical organization and maturation of the
1144 cortex and vascular cylinder inArabidopsis thaliana (Brassicaceae) roots. *Am. J. Bot.* **89**,
1145 908–920 (2002).

1146 53. Cui, H. Middle Cortex Formation in the Root: An Emerging Picture of Integrated
1147 Regulatory Mechanisms. *Mol. Plant* **9**, 771–773 (2016).

1148 54. Ruiz, M., Oustric, J., Santini, J. & Morillon, R. Synthetic Polyploidy in Grafted Crops.
1149 *Front. Plant Sci.* **11**, 540894 (2020).

1150 55. Oh, D.-H. & Dassanayake, M. Landscape of gene transposition-duplication within the
1151 Brassicaceae family. *DNA Res.* **26**, 21–36 (2019).

1152 56. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the
1153 analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).

1154 57. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative
1155 genomics. *Genome Biol.* **20**, 238 (2019).

1156 58. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421
1157 (2009).

1158 59. Omary, M. *et al.* A conserved superlocus regulates above- and belowground root initiation.
1159 *Science* **375**, eabf4368 (2022).

1160 60. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome
1161 alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–
1162 915 (2019).

1163 61. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
1164 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

1165 62. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
1166 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140
1167 (2010).

1168 63. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in
1169 Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329–
1170 337.e4 (2019).

1171 64. Li, S., Yamada, M., Han, X., Ohler, U. & Benfey, P. N. High-Resolution Expression Map
1172 of the Arabidopsis Root Reveals Alternative Splicing and lincRNA Regulation. *Dev. Cell*
1173 **39**, 508–522 (2016).

1174 65. Brady, S. M. *et al.* A high-resolution root spatiotemporal map reveals dominant expression
1175 patterns. *Science* **318**, 801–806 (2007).

1176 66. Vlot, A. H. C., Maghsudi, S. & Ohler, U. Cluster-independent marker feature identification
1177 from single-cell omics data using SEMITONES. *Nucleic Acids Res.* **50**, e107 (2022).

1178 67. Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level
1179 relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).

1180 68. Kumar, L. & E Futschik, M. Mfuzz: a software package for soft clustering of microarray
1181 data. *Bioinformation* **2**, 5–7 (2007).

1182 69. Cardoso-Moreira, M. *et al.* Gene expression across mammalian organ development. *Nature*
1183 **571**, 505–509 (2019).

1184 70. Quigley, F., Rosenberg, J. M., Shachar-Hill, Y. & Bohnert, H. J. From genome to function:
1185 the Arabidopsis aquaporins. *Genome Biol.* **3**, RESEARCH0001 (2002).

1186 71. Wagner, U., Edwards, R., Dixon, D. P. & Mauch, F. Probing the diversity of the
1187 Arabidopsis glutathione S-transferase gene family. *Plant Mol. Biol.* **49**, 515–532 (2002).

1188 72. Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. MAFFT: a novel method for rapid
1189 multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–
1190 3066 (2002).

1191 73. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood
1192 trees for large alignments. *PLoS One* **5**, e9490 (2010).

1193 74. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. Ggtree: An r package for
1194 visualization and annotation of phylogenetic trees with their covariates and other associated
1195 data. *Methods Ecol. Evol.* **8**, 28–36 (2017).

1196 75. Andrijevic, D. *et al.* Cellular recovery after prolonged warm ischaemia of the whole body.
1197 *Nature* **608**, 405–412 (2022).

1198 76. Squair, J. W. *et al.* Confronting false discoveries in single-cell differential expression. *Nat.*
1199 *Commun.* **12**, 5692 (2021).

1200 77. Murphy, A. E. & Skene, N. G. A balanced measure shows superior performance of
1201 pseudobulk methods in single-cell RNA-sequencing analysis. *Nat. Commun.* **13**, 7851
1202 (2022).

1203 78. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor
1204 RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–
1205 4297 (2012).

1206 79. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
1207 for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

1208 80. Gasch, A. P. & Eisen, M. B. Exploring the conditional coregulation of yeast gene

expression through fuzzy k-means clustering. *Genome Biol.* **3**, RESEARCH0059 (2002).

81. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

82. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).

83. Álvarez-Carretero, S., Kapli, P. & Yang, Z. Beginner's Guide on the Use of PAML to Detect Positive Selection. *Mol. Biol. Evol.* **40**, (2023).

84. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449 (2005).

85. Wang, G., Oh, D.-H. & Dassanayake, M. GOMCL: a toolkit to cluster, evaluate, and extract non-redundant associations of Gene Ontology-based functions. *BMC Bioinformatics* **21**, 139 (2020).

86. Kurihara, D., Mizuta, Y., Sato, Y. & Higashiyama, T. ClearSee: a rapid optical clearing reagent for whole-plant fluorescence imaging. *Development* **142**, 4168–4179 (2015).

87. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

88. Bouché, F. Arabidopsis - Roots: cross section. (2019) doi:10.6084/M9.FIGSHARE.4688761.

ACKNOWLEDGEMENTS

This work was supported by the Department of Energy (BER-DE-SC0020358 to J.R.D., J.S., and M.D.; BER DE-SC0022985 to J.R.D and M.D.), the US National Science Foundation (NSF-BSF-IOS-EDGE 1923589/2019610 to M.D.), and Howard Hughes Medical Institute Investigator award (JRD). Single-cell library prep and next-generation sequencing was carried out in the Advanced Genomics Core at the University of Michigan, supported by the National Cancer Institutes of Health Award P30CA046592 to J.S..The authors acknowledge Dr. Song Li and his lab members at Virginia Tech for sharing feedback during group discussions. The authors also acknowledge the LSU High Performance Computing services and Stanford Research Computing Center for providing computational resources needed for data analyses.

AUTHOR CONTRIBUTIONS

Conceptualization, G.W., J.R.D., J.S., and M.D.; Methodology, G.W., K.H.R., J.R.D., J.S., and M.D.; Software, G.W.; Validation, A.D.; Formal Analysis and Investigation, G.W.; Resources, D-H.O., M.O. and R. L.; Data Curation, G.W.; Writing - original draft, G.W. and M.D.; Writing - Review & Editing, all authors; Visualization, G.W.; Supervision, J.R.D., J.S., and M.D.; Funding Acquisition, J.R.D., J.S., and M.D.

1245 **DECLARATION OF INTERESTS**

1246 The authors declare no competing interests.

1247

1248 **FIGURE LEGENDS**

1249 **Fig. 1. Single cell RNA profiling of root cells across 4 diploid Brassicaceae species:**
1250 **Arabidopsis, *Sisymbrium irio*, *Schrenkiella parvula*, and *Eutrema salsugineum*.** a Integrated
1251 UMAP view of root atlas (174,820 cells) from 4 diploid Brassicaceae species representing 10
1252 major root cell types. The integrated atlas included control, 5 μM ABA- and 100 mM NaCl-
1253 treated samples. b Schematic illustration of major Arabidopsis root cell types annotated in
1254 current study. Illustration was adapted from the Plant Illustrations repository[88]. c Root cell atlas
1255 for individual species on the integrated UMAP space. The atlas for each species included control,
1256 5 μM ABA- and 100 mM NaCl-treated samples. d Preferential expression of known cell-type
1257 markers based on cells from all species under control condition. See Supplementary Fig. 8 for
1258 detailed preferential marker gene expression in each of the four species. See Supplementary Data
1259 2 for the list of markers used. Dot size represents the percentage of cells in which each gene is
1260 expressed. Dot color intensity indicates the normalized expression of each gene in each cell type.
1261 Numbers in parenthesis indicate the number of marker genes per cell type. QC, Quiescent center;
1262 LRC, Lateral Root Cap; Col, Columella; Atr, Atrichoblast; Tri, Trichoblast; Cor, Cortex; End,
1263 Endodermis; Per, Pericycle; Pro, Procambium; Phl, Phloem; Xyl, Xylem. e Number of known
1264 Arabidopsis markers that were detected in each species under control condition in the current
1265 study. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Spa, *Schrenkiella parvula*; Sir, *Sisymbrium*
1266 *irio*.

1267 **Fig. 2. Whole-mount *in situ* HCR of cell type markers that are conserved between the 4**
1268 **Brassicaceae species under control condition.** Left panel, the expression of selected cell type
1269 markers in our integrated atlas from the control samples (for visualization purpose only, the
1270 markers were identified separately for each species); middle panel, schematic representation of
1271 Arabidopsis root cell types with the target cell type highlighted in magenta; right panel, whole-
1272 mount *in situ* HCR images of selected cell type markers across species. Note that *S. irio* and *S.*
1273 *parvula* have two layers of cortex. Arrows in yellow mark the cell type where the marker
1274 expression is expected. PRPL1, proline-rich protein-like 1; ELP, extensin-like protein; UCC1,
1275 uclacyanin 1; SIAR1, siliques are red 1; LAX3, like aux1 3; SAPL, Sister of APL; FLA11,
1276 FASCICLIN-like arabinogalactan-protein 11. Scale bars, 100 μm.

1277 **Fig. 3. Evolution of gene expression profiles across cell types under stress-neutral**
1278 **conditions.** a Distribution of Pearson correlation coefficients for the cell type expression of
1279 15,027 1-to-1 orthologs between each pair of species under control condition. The cell type
1280 expression was calculated by averaging the normalized expression level for each gene across all
1281 the single cells within each cell type, see Methods for details. The line and text in red indicate the
1282 median. b The numbers of orthologs showing diverged expression profiles between species
1283 groups under control condition. Orthologs whose cell type expression in one species group is not
1284 conserved in another species group were considered as orthologs with diverged expression

profiles, see details in Methods. c-d Cell type expression of selected examples with diverged expression across species from control samples, *NIP5;1* (c) and *GSTU25* (d). Asterisks indicate where the diverged expression was detected. Left, expected expression of the selected examples based on scRNA-seq data visualized on UMAP space. Right, pseudo-bulk cell type expression of the selected examples. LRC, Lateral Root Cap; Col, Columella; Atr, Atrichoblast; Tri, Trichoblast; Cor, Cortex; End, Endodermis; Per, Pericycle; Pro, Procambium; Phl, Phloem; Xyl, Xylem. e-f Whole-mount *in situ* HCR of *NIP5;1* (e) and *GSTU25* (f) across species under control condition. Arrows in yellow mark the cell type where the diverged expression is expected. Scale bar, 100 µm. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Spa, *Schrenkiella parvula*; Sir, *Sisymbrium irio*.

Fig. 4. Comparison of cortex subpopulations among species from control samples. a UMAP representation of three cortex subpopulations in *S. irio*. b Projection of cortex cells from Arabidopsis, *E. salsugineum*, and *S. parvula* onto the three cortex subpopulations in *S. irio*. c Proportion of the three cortex subpopulations in the four Brassicaceae species. Asterisks represent significance determined by comparing each species to *S. irio* using hypergeometric test at p-value ≤ 0.01 with red asterisks indicating over-representation and black asterisks indicating under-representation. d Normalized expression of *Si-s1384-01650* (OG14187) on UMAP space (top) and its whole-mount *in situ* HCR (bottom) in *S. irio*. Arrows in yellow mark the cortex layers. Scale bars, 100 µm. e Normalized expression of *Si-s1384-01650* orthologs (OG14187) on UMAP space (top) and their whole-mount *in situ* HCR (bottom) in Arabidopsis, *E. salsugineum*, and *S. parvula*. Arrows in yellow mark the cortex layers. Scale bars, 100 µm. f Normalized expression of genes involved in flavonoid biosynthesis in inner and outer cortex layers in *S. irio* (left) and *S. parvula* (right). Red dots indicate genes that were significantly differentially expressed (adjusted p-value ≤ 0.01 in FindMarkers) between the two cortex layers and size of the dots corresponds to the $\log_2$FC. PAL1/2, phenylalanine ammonia-lyase1/2; C4H, cinnamic acid 4-hydroxylase; 4CL3, 4-coumaric acid:CoA ligase; ACC1, acetyl-CoA carboxylase; CHS, chalcone synthase; CHI, chalcone isomerase; CHIL, chalcone isomerase-like; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; FLS1/3, flavonol synthase1/3; DFR, dihydroflavonol reductase; ANS, anthocyanin synthase; ANR, anthocyanin reductase; LAC15, laccase 15; FNSI, flavone synthase I.

Fig. 5. Divergent cell type transcriptomic response to ABA and NaCl. a Principal component (PC) analysis of pseudo-bulk cell-type transcriptomes from control, 5 µM ABA- and 100 mM NaCl-treated samples. Each symbol represents a pseudo-bulk cell-type transcriptome. The color and the shape of each symbol correspond to the cell type and the condition, respectively. Samples from each species are encircled by a line (based on 95% confidence level) and labeled. The red dashed lines separate cell type transcriptome subgroups within each species. b Comparison of transcriptome responsiveness as averaged Augur area under the curve (AUC) scores across cell types in response to ABA and NaCl. AUC here quantifies the separability between control and treatment samples for each cell type. AUC = 0.5 indicates random separability; AUC = 1.0 indicates perfect separability. Asterisks indicate significance after Wilcoxon testing (n = 50) at adjusted p-value cutoff of 0.05. c Hierarchical clustering of species based on Euclidean distance between their cell type transcriptome responses in (b). d Summary

of GO functional enrichment among DEGs in response to ABA (left) and NaCl (right) for each cell type in each species. The detailed list of GO terms is present in Supplementary Data 7. The colors of the tiles represent different cell types. The shade represents the *p*-value assigned by the enrichment test at adjusted *p*-value < 0.05 with darker shades indicating smaller *p*-values. e Principal component analysis of transcriptomic divergence (as AUC scores) between species for each cell type. AUC here quantifies the separability between species for each cell type and was calculated for each pair of species under control, ABA and NaCl treatments separately. Each symbol represents the transcriptomic divergence between six possible species pairs under one condition. The color and the shape of each symbol correspond to the cell type and the condition, respectively. f Transcriptomic distinctiveness (as AUC scores) between species pairs under different conditions. AUC here quantifies the separability between species for each cell type. AUC = 0.5 indicates random separability; AUC = 1.0 indicates perfect separability. Asterisks indicate significance after comparing to control (n = 50) with Student's t-test at *p*-value cutoff of 0.05. The red dash lines mark the medians the transcriptomic distinctiveness under control condition. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Sir, *Sisymbrium irio;* Spa, *Schrenkiella parvula*.

Fig. 6. Stress responses between orthologs. a-b Differences between observed and bootstrapped (10,000 times) occurrences of response patterns among all 1-to-1 orthologs (a) and orthologs showing convergent response (b). The darker and lighter color intensities represent significant and insignificant deviation of the observed occurrences from the expected distribution, respectively, at Benjamini-Hochberg-adjusted *p*-value ≤ 0.01. c Percentage of divergently regulated orthologs (in response to 5 μM ABA) with conserved or divergent ABF binding. ABF binding was defined as the presence of ABF binding events in the 2-kilobase upstream of the transcription start site using the DAP-seq data from Sun *et al.*, 2022[9]. See Supplementary Fig. 33a for more explanations. Numbers on the top represent *p*-values determined by the Wilcoxon test. Asterisks represent significance at the *p*-value cutoff of 0.05. d Contributions of each cell type to co-expression networks. Cluster ID and the number of orthologs in the cluster are shown in parentheses in the Y-axis labels. Each gray line represents the expression of an ortholog across conditions and species. The red line indicates the median of all orthologs in the given cluster. Cell type proportions of the cluster were compared to background (bg) where orthologs from all clusters were averaged (see Methods for details). Asterisks represent significance determined by hypergeometric test, * *p*-value < 0.01, ** *p*-value < 0.001, *** *p*-value < 0.0001. LRC, Lateral Root Cap; Col, Columella; Atr, Atrichoblast; Tri, Trichoblast; Cor, Cortex; End, Endodermis; Per, Pericycle; Pro, Procambium; Phl, Phloem; Xyl, Xylem. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Sir, *Sisymbrium irio;* Spa, *Schrenkiella parvula*.

Fig. 7. Cell type stress responses in *C. sativa*. a Root single cell atlas (27,442 cells) for *C. sativa*. The atlas included control, 5 μM ABA- and 100 mM NaCl-treated samples. b Principal component analysis of pseudo-bulk cell-type transcriptomes. Each symbol represents a pseudo-bulk cell-type transcriptome from control, 5 μM ABA- and 100 mM NaCl-treated samples. The color and the shape of each symbol correspond to the cell type and the condition, respectively. c Comparison of transcriptome responsiveness as averaged Augur area under the curve (AUC)

scores across cell types in response to ABA and NaCl. AUC here quantifies the separability between control and treatment samples for each cell type. AUC = 0.5 indicates random separability; AUC = 1.0 indicates perfect separability. Asterisks indicate significance after Wilcoxon testing (n = 50) at adjusted *p*-value cutoff of 0.05. d Summary of GO functional enrichment among DEGs combined from all sub-genomes in response to ABA (top) and NaCl (bottom) for each cell type in *C. sativa*. The detailed list of GO terms is present in Supplementary Data 7. The shade represents the *p*-value assigned by the enrichment test at adjusted *p*-value < 0.05 with darker shades indicating smaller *p*-values. e Comparison of transcriptome responsiveness (as AUC scores) for each sub-genome in *C. sativa*. AUC here quantifies the separability between control and treatment samples using each sub-genome for each cell type. AUC = 0.5 indicates random separability; AUC = 1.0 indicates perfect separability. Lowercase letters on top of violin plots indicate significantly different groups (n = 50) as determined by one-way analysis of variance (ANOVA) with post-hoc Tukey honest significant difference (HSD) testing at *p*-value < 0.01. f Ranked homoeologous triplets based on their divergence in coding sequence (①), expression (②), cell type specificity (③) as well as expression changes in response to stresses (④ and ⑤). Numbers on the left indicate the number of candidates considered for each category. Endodermis, phloem and xylem cells were removed from the comparison for NaCl treatment (n.d.) due to insufficient number of cells (< 50 cells) in c-f. N/A, not applicable.

1404 **Supplemental information**

1405 **Supplementary Figures 1-39, Supplementary Data 1-10, and Supplementary Movie 1**

1406 **Supplementary Fig. 1. Overview of scRNA-seq dataset for four 5 Brassicaceae species and**
1407 **data preprocessing.** The number in each UMAP represents the total number of filtered cells
1408 from control, 5 μM ABA- and 100 mM NaCl-treated samples.

1409 **Supplementary Fig. 2. Percentage of genes with orthologs between Arabidopsis and other**
1410 **species**. Csa, *Camelina sativa*; Esa, *Eutrema salsugineum*; Spa, *Schrenkiella parvula*; Sir,
1411 *Sisymbrium irio*; Mtr, *Medicago truncatula*; Sly, *Solanum lycopersicum*; Osa, *Oryza sativa*;
1412 Zma, *Zea mays*. Numbers in parenthesis indicate the numbers of protein-coding genes for each
1413 species.

1414 **Supplementary Fig. 3. Knee plots for each sample in our study.** For each species, samples
1415 were harvested from control, 5μM ABA-, and 100 mM NaCl-treated conditions. Two
1416 independent replicates (R1 and R2) were used for each condition. The high-confidence cell-
1417 associated barcodes were clearly separated from the background ones.

1418 **Supplementary Fig. 4. A two-step strategy for cross-species cell type annotations using**
1419 **three independent methods.** The Arabidopsis cell type reference transcriptomes were from Li
1420 *et al.*, 2016[64] and Brady *et al.*, 2007[65]. The different sets of highly-variable genes were obtained
1421 by categorizing them into different percentile bins. The key parameters for each method were
1422 selected based on the tools used, see details in Methods.

1423 **Supplementary Fig. 5. Root cell atlas for each species under different conditions on the**
1424 **integrated UMAP space.** All ten root cell types were detected in each species under different
1425 conditions. LRC, Lateral Root Cap.

1426 **Supplementary Fig. 6. Developmental zone annotation for the integrated root cell atlas**
1427 **from four species (top) and in each species (bottom)**. All developmental zones were detected
1428 in each species under different conditions. LRC, Lateral Root Cap.

1429 **Supplementary Fig. 7. Projection of Arabidopsis control cells from current study to Shahan**
1430 ***et al*[21]. The red dashed oval marks the expected space for quiescent center cells, currently not
1431 detected in the current dataset for Arabidopsis.

1432 **Supplementary Fig. 8. Cell-type expression of known Arabidopsis cell-type markers and**
1433 **their 1-to-1 ortholog in other species under control condition.** See Supplementary Data 2 for
1434 the list of markers used. Dot size represents the percentage of cells in which each gene is
1435 expressed. Dot color intensity indicates the normalized expression of each gene in each cell type.
1436 QC, Quiescent center; LRC, Lateral Root Cap; Col, Columella; Atr, Atrichoblast; Tri,
1437 Trichoblast; Cor, Cortex; End, Endodermis; Per, Pericycle; Pro, Procambium; Phl, Phloem; Xyl,
1438 Xylem.

1439 **Supplementary Fig. 9. Expression of conserved (a) and diverged (b) Arabidopsis cell type**
1440 **markers across species under control condition.** Red boxes mark the species where these cell

1441 type markers lost their expression specificity. Tiles with incomplete fill indicate these genes were
1442 expressed in less than 10% of cells from the target cell types. They were not considered for
1443 marker identification.

1444 **Supplementary Fig. 10. Identification of cell type preferentially enriched genes (PEGs) that**
1445 **were conserved across species for each cell type in the current study.** Ath, Arabidopsis; Esa,
1446 *Eutrema salsugineum*; Spa, *Schrenkiella parvula*; Sir, *Sisymbrium irio*. The conserved PEGs for
1447 each cell type are highlighted with red boxes.

1448 **Supplementary Fig. 11. Expression of selected cell type markers in Fig. 2 across species.**
1449 The expression shown is from the control samples. PRPL1, proline-rich protein-like 1; ELP,
1450 extensin-like protein; UCC1, uclacyanin 1; SIAR1, siliques are red 1; LAX3, like aux1 3; SAPL,
1451 Sister of APL; FLA11, FASCICLIN-like arabinogalactan-protein 11.

1452 **Supplementary Fig. 12. Comparison of the average and the standard deviation of cell type**
1453 **specificity index for the 1-to-1 orthologs.** Cell type specificity index (τ) is based on tau metric
1454 of tissue specificity in the control samples of each species and calculated as
1455 $\sum_{i=1}^{N}(1-x_i)/(N-1)$, where N is the number of cell types and x_i is the expression level in cell
1456 type is normalized by the maximal expression of the given gene across all cell types. Cell type
1457 specificity index (τ) ranges from 0 to 1, where 0 means broad expression in all cell types and 1
1458 indicates specific expression in one cell type. The cell type specificity indexes for each 1-to-1
1459 orthologs were used to calculate the average and standard deviation for these orthologs.

1460 **Supplementary Fig. 13. Phylogenetic tree of NIPs (a) and GSTUs (b) from Arabidopsis, *S.***
1461 ***irio*, *S. parvula*, and *E. salsugineum*.** Genes marked with asterisks are the reciprocal best
1462 matches of each other.

1463 **Supplementary Fig. 14. Expression of *SpGSTU25* (*Sp1g15320*) (a) and *Sp1g15325* (b) based**
1464 **on scRNA-seq data for *S. parvula* under control condition.** Left panels display the expression
1465 of the selected genes visualized on UMAP space. Right panels show the pseudo-bulk cell type
1466 expression of the selected genes. LRC, Lateral Root Cap; Col, Columella; Atr, Atrichoblast; Tri,
1467 Trichoblast; Cor, Cortex; End, Endodermis; Per, Pericycle; Pro, Procambium; Phl, Phloem; Xyl,
1468 Xylem.

1469 **Supplementary Fig. 15. Comparison of cortex cells across species from control samples.** a
1470 Cross-sections of roots at the maturation zone in each species. Roots were obtained from 5–6-
1471 day old seedlings for each species. Scale bars, 20 μm. b UMAP representation of cortex cells in
1472 each species. Sub-populations in each species are manually circled based on trajectories
1473 identified by Monocle.

1474 **Supplementary Fig. 16. Functional representations of genes that were preferentially**
1475 **expressed (a) and depleted (b) in the middle cortex subpopulation of *S. irio*.** Numbers in the
1476 parenthesis on the y-axis indicate the number of genes. Box plots represent the median (central
1477 value), upper and lower quartile (box limits) with whiskers at ± 1.5 × the interquartile range. The
1478 red dash line in the line graph in (a) marks the normalized expression of *Si-s1384-01650*, the

1479 HCR result of its expression in *S. irio* is shown in Fig. 4d. GO clusters are differently colored
1480 and labeled with the representative functional term. Each node represents a gene ontology (GO)
1481 term; node size represents genes in the test set assigned to that functional term; GO terms
1482 sharing > 50% of genes are connected with edges; and the shade of each node represents the *p*-
1483 value assigned by the enrichment test at adjusted *p* values < 0.05 with darker shades indicating
1484 smaller *p*-values.

1485 **Supplementary Fig. 17. Expression of genes that are preferentially expressed in inner (top)**
1486 **and outer (bottom) cortex layers in *S. irio* and their functional representations.** GO clusters
1487 are differently colored and labeled with the representative functional term. Each node represents
1488 a gene ontology (GO) term; node size represents genes in the test set assigned to that functional
1489 term; GO terms sharing > 50% of genes are connected with edges; and the shade of each node
1490 represents the p-value assigned by the enrichment test at adjusted p values < 0.05 with darker
1491 shades indicating smaller p-values.

1492 **Supplementary Fig. 18. The expression of flavonoid biosynthetic genes in rice and**
1493 **Medicago.** The spatial separation of the flavonoid pathway in different cortex layers is not
1494 conserved in rice and Medicago. C4H, cinnamic acid 4-hydroxylase; CHS, chalcone synthase;
1495 CHIL, chalcone isomerase-like.

1496 **Supplementary Fig. 19. Principal component (PC) analysis of pseudo-bulk cell-type**
1497 **transcriptomes generated from single-nucleus RNA-seq of Arabidopsis (Farmer *et al.*,**
1498 **2021[37]).** Each symbol represents a pseudo-bulk cell-type transcriptome. The red dashed lines
1499 separate cell type transcriptome subgroups within each species.

1500 **Supplementary Fig. 20. Expression changes of OG04736 orthologs (*AT3G15670*) across**
1501 **species in response to ABA.** a Expression of OG04736 orthologs across species from scRNA-
1502 seq data under different conditions. Left, UMAP representation of the expression of OG04736
1503 orthologs. Right, pseudo-bulk cell type expression of OG04736 orthologs. b Whole-mount *in situ*
1504 HCR detection of expression changes of OG04736 orthologs across species in response to ABA.
1505 Left, confocal images of whole-mount HCR showing the detection of OG04736 orthologs across
1506 species. Scale bars, 150 μm. Right, average signal intensity of OG04736 orthologs across defined
1507 cell type groups.

1508 **Supplementary Fig. 21. Expression changes of OG03053 orthologs (*AT2G37870*) across**
1509 **species in response to ABA.** a Expression of OG03053 orthologs across species from scRNA-
1510 seq data under different conditions. Left, UMAP representation of the expression of OG03053
1511 orthologs. Right, pseudo-bulk cell type expression of OG03053 orthologs. b Whole-mount *in situ*
1512 HCR detection of expression changes of OG03053 orthologs across species in response to ABA.
1513 Left, confocal images of whole-mount HCR showing the detection of OG03053 orthologs across
1514 species. Scale bars, 50 μm. Right, average signal intensity of OG03053 orthologs across defined
1515 cell type groups.

1516 **Supplementary Fig. 22. Expression changes of OG10873 orthologs (*AT5G47450*) across**
1517 **species in response to NaCl.** a Expression of OG10873 orthologs across species from scRNA-

1518 seq data under different conditions. Left, UMAP representation of the expression of OG10873
1519 orthologs. Right, pseudo-bulk cell type expression of OG10873 orthologs. b Whole-mount *in situ*
1520 HCR detection of expression changes of OG10873 orthologs across species in response to ABA.
1521 Left, confocal images of whole-mount HCR showing the detection of OG10873 orthologs across
1522 species. Scale bars, 50 µm. Right, average signal intensity of OG10873 orthologs across defined
1523 cell type groups.

1524 **Supplementary Fig. 23. Number of stress responsive genes in each cell type from each**
1525 **species in response to 5 µM ABA and 100 mM NaCl.** Total number of stress responsive genes
1526 (nr, non-redundant) from all cell types in each species was shown on the top left corner.

1527 **Supplementary Fig. 24. Overlap of stress responsive genes in each developmental zone**
1528 **from each species in response to 5 µM ABA (left) and 100 mM NaCl (right).**

1529 **Supplementary Fig. 25. The percentage of DEGs shared between ABA and NaCl treatment.**
1530 Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Sir, *Sisymbrium irio*; Spa, *Schrenkiella parvula*.

1531 **Supplementary Fig. 26. Euclidean distance between treated and control samples in Fig. 5e.**
1532 The salt-treated samples were generally more distant (greater Euclidean distances) to the control
1533 samples than the ABA-treated samples.

1534 **Supplementary Fig. 27. Transcriptomic distinctiveness between species pairs under**
1535 **different conditions**. AUC here quantifies the separability between species for each cell type.
1536 AUC = 0.5 indicates random separability; AUC = 1.0 indicates perfect separability. Asterisks
1537 indicate significance after comparing to control (n = 50) with Student's t-test at *p*-value cutoff of
1538 0.05. The red dash lines mark the medians the transcriptomic distinctiveness under control
1539 condition. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Sir, *Sisymbrium irio;* Spa, *Schrenkiella*
1540 *parvula*.

1541 **Supplementary Fig. 28. Changes in cell type transcriptome divergence between species in**
1542 **response to ABA (top) and NaCl (bottom).** The darker colors indicate larger differences in
1543 transcriptome divergence induced by the treatment. Asterisks indicate significance after
1544 comparing to control (n = 50) with Student's t-test at p-value cutoff of 0.05. The black boxes
1545 highlight the comparisons of interest. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Sir,
1546 *Sisymbrium irio*; Spa, *Schrenkiella parvula*.

1547 **Supplementary Fig. 29. Differences between observed and bootstrapped (10,000 times)**
1548 **occurrences of different ortholog response patterns.** The patterns were grouped according to
1549 orthologs showing a, lineage-specific gain of response; b, lineage-specific loss of response; c.
1550 synapomorphic response; d. convergent response; e, diverse response; f, conserved response.
1551 Number of orthologs for each pattern in each cell type is indicated in the corresponding tile. The
1552 differences were measured as the number of standard deviations between the overserved
1553 occurrences and the corresponding bootstrapped means and indicated with the shade of the color
1554 with darker colors corresponding to larger differences. Black and grey text represent significant

1555 and insignificant deviation of the observed occurrences from the expected distribution,
1556 respectively, at Benjamini-Hochberg-adjusted *p*-value ≤ 0.01.

1557 **Supplementary Fig. 30. Differences between observed and bootstrapped (10,000 times)**
1558 **occurrences of response patterns among all ortholog groups that have been duplicated only**
1559 **in one species.** The darker and lighter color intensities represent significant and insignificant
1560 deviation of the observed occurrences from the expected distribution, respectively, at Benjamini-
1561 Hochberg-adjusted p-value ≤ 0.01.

1562 **Supplementary Fig. 31. Percentage of genes with ABF binding among genes induced by**
1563 **5μM ABA (left) or 100mM NaCl (right).** ABF binding was defined as the presence of ABF
1564 binding events in the 2-kilobase upstream of the transcription start site using the DAP-seq data
1565 from Sun *et al.*, 2022[9].

1566 **Supplementary Fig. 32. Number of responsive cell types for genes with different numbers**
1567 **of ABF binding sites.** ABF binding was defined as the presence of ABF binding events in the 2-
1568 kilobase upstream of the transcription start site using the DAP-seq data from Sun *et al.*, 2022[9].
1569 Significance was represented by *p*-values determined by Student's t-test and shown on the top of
1570 each comparison. Numbers above each box indicate the number of genes included in each group.

1571 **Supplementary Fig. 33. The relationship between ABF binding and stress responses of**
1572 **orthologs between species.** a Categorization of orthologs with different ABF binding events.
1573 ABF binding was defined as the presence of ABF binding events in the 2-kilobase upstream of
1574 the transcription start site using the DAP-seq data from Sun *et al.*, 2022[9]. b Percentage of
1575 divergently regulated orthologs (in response to 100 mM NaCl) with conserved or divergent ABF
1576 binding. Significance was represented by *p*-values determined by Wilcoxon test and shown on
1577 the top of each comparison.

1578 **Supplementary Fig. 34. Contributions of each cell type to co-expression networks identified**
1579 **from stress-responsive orthologs under 5 μM ABA (a) and 100 mM NaCl (b).** Cluster ID and
1580 the number of orthologs in the cluster are shown in parentheses in the Y-axis labels. Cell type
1581 proportions of the cluster were compared to background (bg) where orthologs from all clusters
1582 were averaged (see Methods for details). Asterisks represent significance determined by
1583 hypergeometric test, * *p*-value < 0.01, ** *p*-value < 0.001, *** *p*-value < 0.0001. LRC, Lateral
1584 Root Cap; Col, Columella; Atr, Atrichoblast; Tri, Trichoblast; Cor, Cortex; End, Endodermis;
1585 Per, Pericycle; Pro, Procambium; Phl, Phloem; Xyl, Xylem. Ath, Arabidopsis; Esa, *Eutrema*
1586 *salsugineum*; Sir, *Sisymbrium irio;* Spa, *Schrenkiella parvula*.

1587 **Supplementary Fig. 35. Functional representations of co-expressed orthologs in response to**
1588 **NaCl.** GO clusters are differently colored and labelled with the representative functional term.
1589 Each node represents a gene ontology (GO) term; node size represents genes in the test set
1590 assigned to that functional term; GO terms sharing > 50% of genes are connected with edges;
1591 and the shade of each node represents the *p*-value assigned by the enrichment test at adjusted *p*
1592 values < 0.05 with darker shades indicating smaller *p*-values.

1593 **Supplementary Fig. 36. Integration between Arabidopsis and *C. sativa* using different sub-**
1594 **genomes in *C. sativa*.**

1595 **Supplementary Fig. 37. Overall transcriptome expression levels separated by the three sub-**
1596 **genomes for each cell type under different conditions.** The number of genes considered for
1597 sub-genome G1, G2, and G3 is 21,848, 21,682, and 23,066, respectively. Numbers in red
1598 indicate the averages of transcript levels.

1599 **Supplementary Fig. 38. Comparison of GO functional enrichment among DEGs in**
1600 **response to ABA (top) and NaCl (bottom) for each cell type across species.** The enrichment
1601 in *C. sativa* was performed on DEGs combined from all sub-genomes. Endodermis, phloem and
1602 xylem cells in *C. sativa* were removed from the comparison for NaCl treatment due to
1603 insufficient number of cells (< 50 cells). The detailed list of GO terms is present in
1604 Supplementary Data 6. The shade represents the *p*-value assigned by the enrichment test at
1605 adjusted *p*-value < 0.05 with darker shades indicating smaller *p*-values.

1606 **Supplementary Fig. 39. Percentage of DEGs in each sub-genome of *C. sativa*.** The percentage
1607 for each genome was calculated considering all detected genes within individual sub-genomes
1608 (a) and only the complete homeolog triplets (b). Numbers on top of the bars represent adjusted *p*-
1609 values from hypergeometric test.

1610 **Supplementary Data 1. Statistics of scRNA-seq data generated in the current study.**

1611 **Supplementary Data 2. Known Arabidopsis markers and their expression in the current**
1612 **study.**

1613 **Supplementary Data 3. Markers identified from the current study.**

1614 **Supplementary Data 4. Orthologs with divergent expression across species.**

1615 **Supplementary Data 5. Genes preferentially expressed in different cortex subpopulations**
1616 **in *S. irio*.**

1617 **Supplementary Data 6. Differentially expressed genes from each species in response to**
1618 **ABA and NaCl.**

1619 **Supplementary Data 7. GOs enriched among stress-responsive genes in each cell type from**
1620 **each species.**

1621 **Supplementary Data 8. Stress response patterns among orthologs.**

1622 **Supplementary Data 9. Homeologs exhibiting high divergence in coding sequences and**
1623 **expression.**

1624 **Supplementary Data 10. Comparison of gene expression between bulk and pseudo-bulk**
1625 **samples.**

1626 **Supplementary Movie 1. 3D view of integrated atlas.**

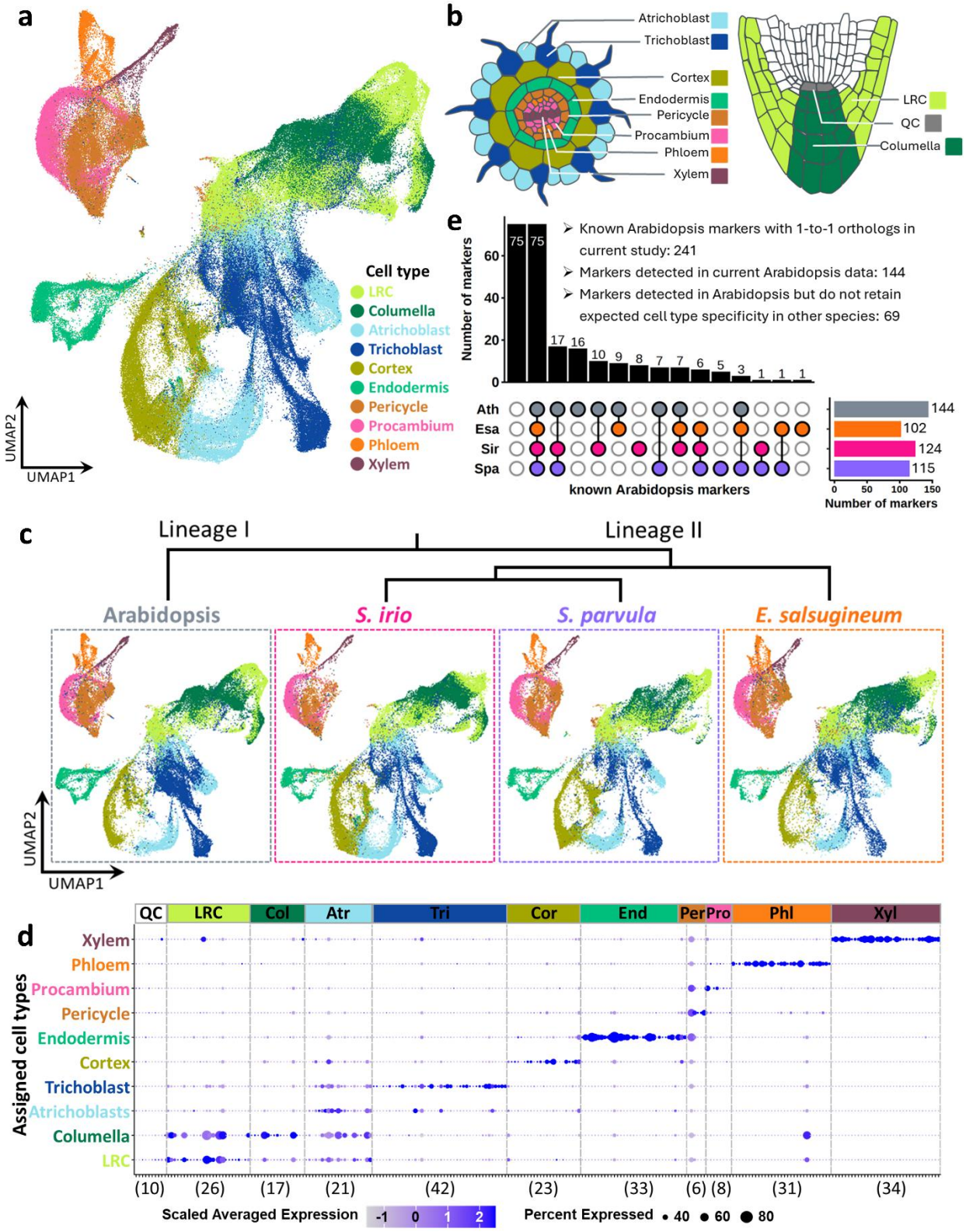


Fig. 1. Single cell RNA profiling of root cells across 4 diploid Brassicaceae species: Arabidopsis, *Sisymbrium irio*, *Schrenkiella parvula*, and *Eutrema salsugineum*. a Integrated

UMAP view of root atlas (174,820 cells) from 4 diploid Brassicaceae species representing 10 major root cell types. The integrated atlas included control, 5 μM ABA- and 100 mM NaCl-treated samples. b Schematic illustration of major Arabidopsis root cell types annotated in current study. Illustration was adapted from the Plant Illustrations repository[88]. c Root cell atlas for individual species on the integrated UMAP space. The atlas for each species included control, 5 μM ABA- and 100 mM NaCl-treated samples. d Preferential expression of known cell-type markers based on cells from all species under control condition. See Supplementary Fig. 8 for detailed preferential marker gene expression in each of the four species. See Supplementary Data 2 for the list of markers used. Dot size represents the percentage of cells in which each gene is expressed. Dot color intensity indicates the normalized expression of each gene in each cell type. Numbers in parenthesis indicate the number of marker genes per cell type. QC, Quiescent center; LRC, Lateral Root Cap; Col, Columella; Atr, Atrichoblast; Tri, Trichoblast; Cor, Cortex; End, Endodermis; Per, Pericycle; Pro, Procambium; Phl, Phloem; Xyl, Xylem. e Number of known Arabidopsis markers that were detected in each species under control condition in the current study. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Spa, *Schrenkiella parvula*; Sir, *Sisymbrium irio*.

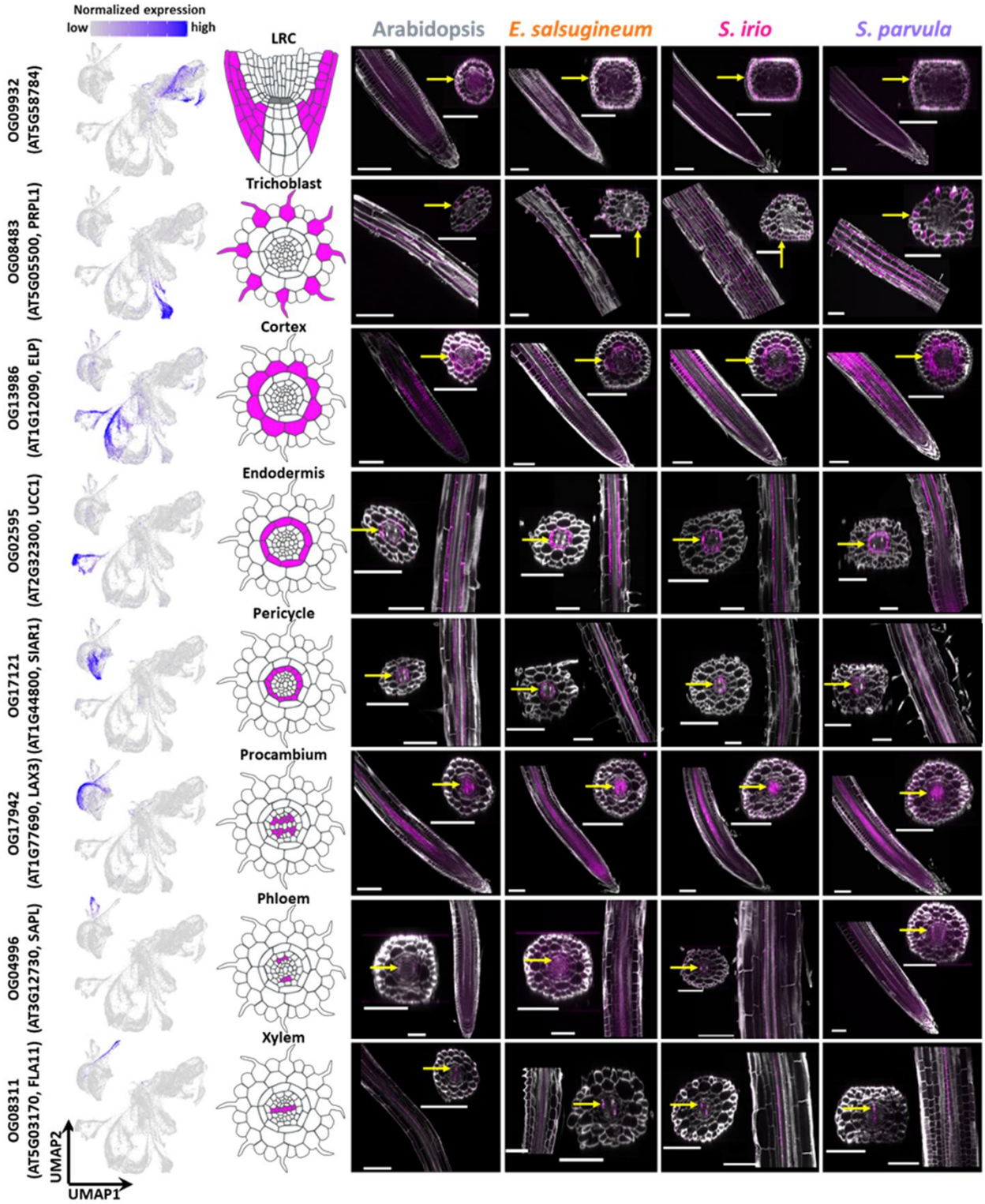


Fig. 2. Whole-mount *in situ* HCR of cell type markers that are conserved between the 4 Brassicaceae species under control condition. Left panel, the expression of selected cell type markers in our integrated atlas from the control samples (for visualization purpose only, the markers were identified separately for each species); middle panel, schematic representation of

Arabidopsis root cell types with the target cell type highlighted in magenta; right panel, whole-mount *in situ* HCR images of selected cell type markers across species. Note that *S. irio* and *S. parvula* have two layers of cortex. Arrows in yellow mark the cell type where the marker expression is expected. PRPL1, proline-rich protein-like 1; ELP, extensin-like protein; UCC1, uclacyanin 1; SIAR1, siliques are red 1; LAX3, like aux1 3; SAPL, Sister of APL; FLA11, FASCICLIN-like arabinogalactan-protein 11. Scale bars, 100 μm.

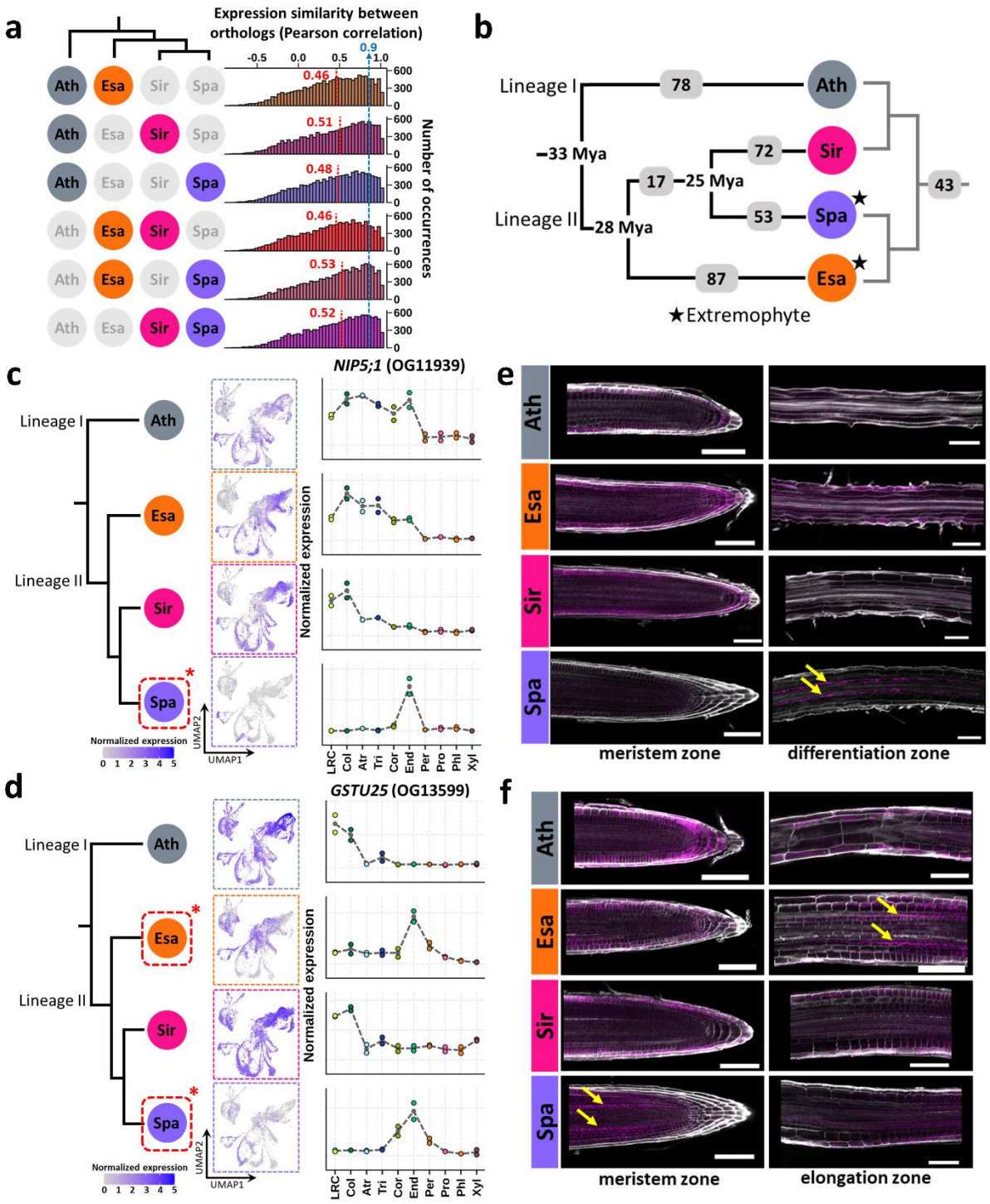


Fig. 3. Evolution of gene expression profiles across cell types under stress-neutral conditions. a Distribution of Pearson correlation coefficients for the cell type expression of 15,027 1-to-1 orthologs between each pair of species under control condition. The cell type expression was calculated by averaging the normalized expression level for each gene across all the single cells within each cell type, see Methods for details. The line and text in red indicate the

median. b The numbers of orthologs showing diverged expression profiles between species groups under control condition. Orthologs whose cell type expression in one species group is not conserved in another species group were considered as orthologs with diverged expression profiles, see details in Methods. c-d Cell type expression of selected examples with diverged expression across species from control samples, *NIP5;1* (c) and *GSTU25* (d). Asterisks indicate where the diverged expression was detected. Left, expected expression of the selected examples based on scRNA-seq data visualized on UMAP space. Right, pseudo-bulk cell type expression of the selected examples. LRC, Lateral Root Cap; Col, Columella; Atr, Atrichoblast; Tri, Trichoblast; Cor, Cortex; End, Endodermis; Per, Pericycle; Pro, Procambium; Phl, Phloem; Xyl, Xylem. e-f Whole-mount *in situ* HCR of *NIP5;1* (e) and *GSTU25* (f) across species under control condition. Arrows in yellow mark the cell type where the diverged expression is expected. Scale bar, 100 µm. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Spa, *Schrenkiella parvula*; Sir, *Sisymbrium irio*.

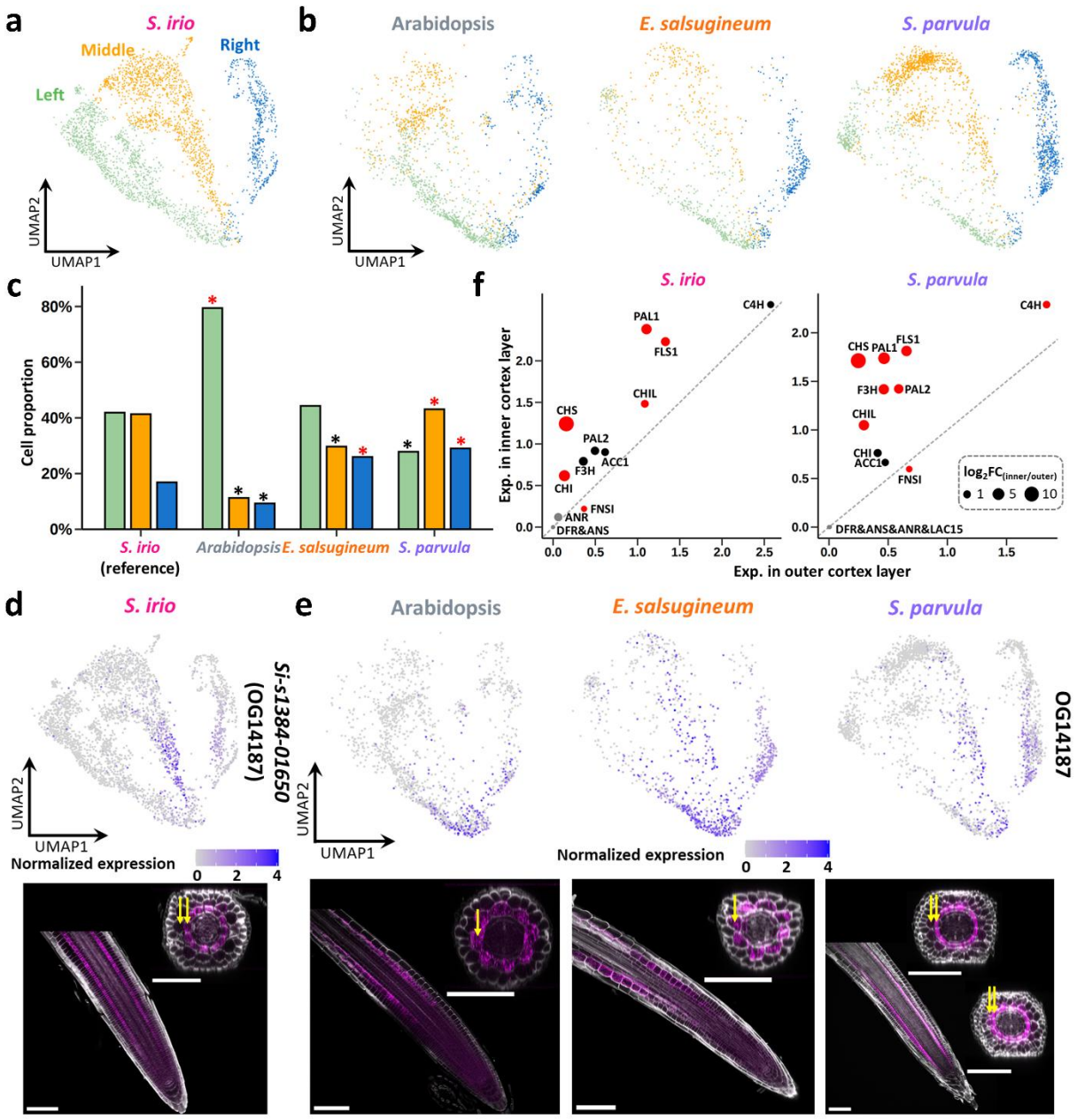


Fig. 4. Comparison of cortex subpopulations among species from control samples. a UMAP representation of three cortex subpopulations in *S. irio*. b Projection of cortex cells from Arabidopsis, *E. salsugineum*, and *S. parvula* onto the three cortex subpopulations in *S. irio*. c Proportion of the three cortex subpopulations in the four Brassicaceae species. Asterisks represent significance determined by comparing each species to *S. irio* using hypergeometric test at *p*-value ≤ 0.01 with red asterisks indicating over-representation and black asterisks indicating under-representation. d Normalized expression of *Si-s1384-01650* (OG14187) on UMAP space (top) and its whole-mount *in situ* HCR (bottom) in *S. irio*. Arrows in yellow mark the cortex layers. Scale bars, 100 µm. e Normalized expression of *Si-s1384-01650* orthologs (OG14187) on UMAP space (top) and their whole-mount *in situ* HCR (bottom) in Arabidopsis, *E. salsugineum*,

and *S. parvula*. Arrows in yellow mark the cortex layers. Scale bars, 100 µm. f Normalized expression of genes involved in flavonoid biosynthesis in inner and outer cortex layers in *S. irio* (left) and *S. parvula* (right). Red dots indicate genes that were significantly differentially expressed (adjusted p-value ≤ 0.01 in FindMarkers) between the two cortex layers and size of the dots corresponds to the $\log_2$FC. PAL1/2, phenylalanine ammonia-lyase1/2; C4H, cinnamic acid 4-hydroxylase; 4CL3, 4-coumaric acid:CoA ligase; ACC1, acetyl-CoA carboxylase; CHS, chalcone synthase; CHI, chalcone isomerase; CHIL, chalcone isomerase-like; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3′-hydroxylase; FLS1/3, flavonol synthase1/3; DFR, dihydroflavonol reductase; ANS, anthocyanin synthase; ANR, anthocyanin reductase; LAC15, laccase 15; FNSI, flavone synthase I.

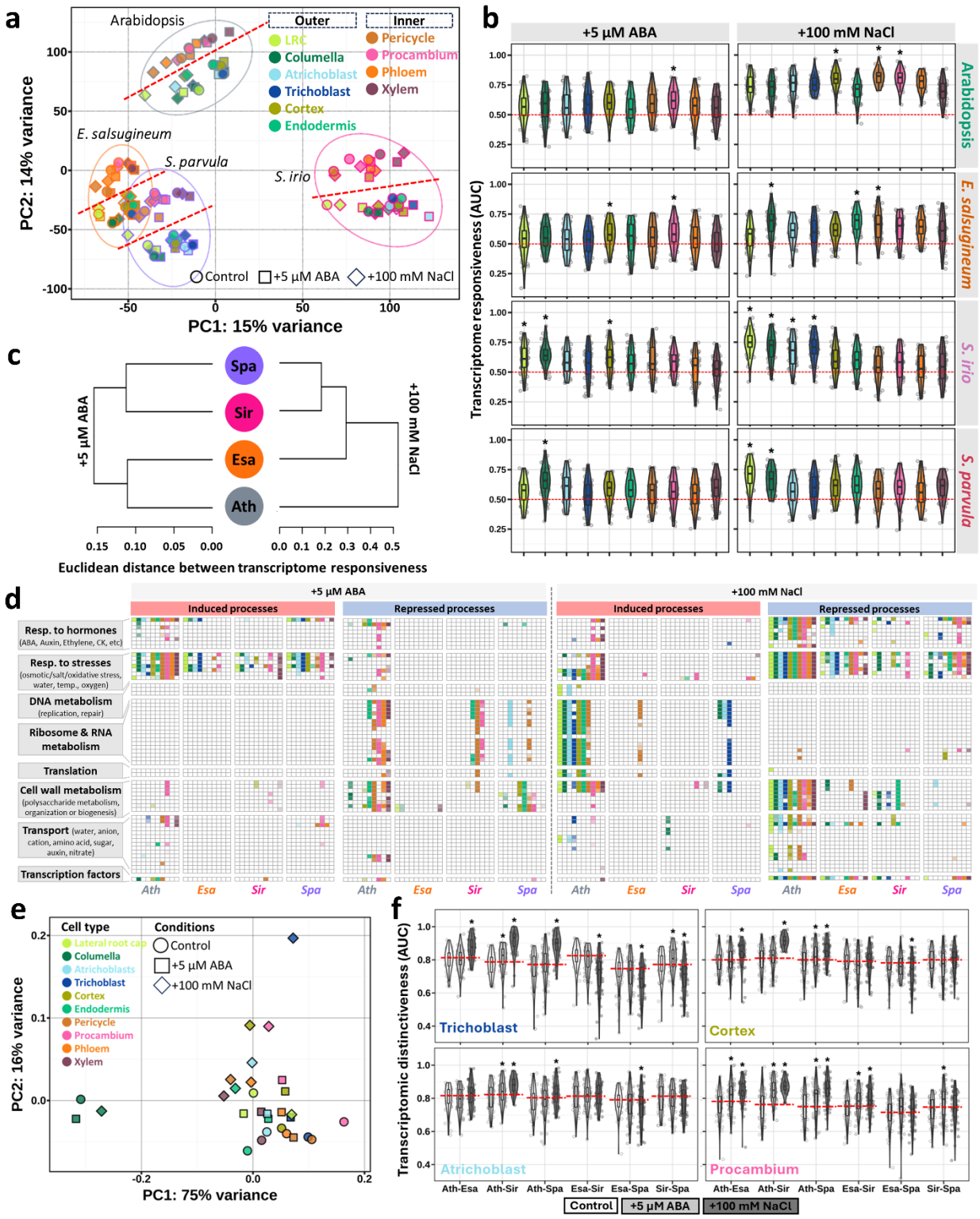


Fig. 5. Divergent cell type transcriptomic response to ABA and NaCl. a Principal component (PC) analysis of pseudo-bulk cell-type transcriptomes from control, 5 µM ABA- and 100 mM NaCl-treated samples. Each symbol represents a pseudo-bulk cell-type transcriptome. The color

and the shape of each symbol correspond to the cell type and the condition, respectively. Samples from each species are encircled by a line (based on 95% confidence level) and labeled. The red dashed lines separate cell type transcriptome subgroups within each species. b Comparison of transcriptome responsiveness as averaged Augur area under the curve (AUC) scores across cell types in response to ABA and NaCl. AUC here quantifies the separability between control and treatment samples for each cell type. AUC = 0.5 indicates random separability; AUC = 1.0 indicates perfect separability. Asterisks indicate significance after Wilcoxon testing (n = 50) at adjusted *p*-value cutoff of 0.05. c Hierarchical clustering of species based on Euclidean distance between their cell type transcriptome responses in (b). d Summary of GO functional enrichment among DEGs in response to ABA (left) and NaCl (right) for each cell type in each species. The detailed list of GO terms is present in Supplementary Data 7. The colors of the tiles represent different cell types. The shade represents the *p*-value assigned by the enrichment test at adjusted *p*-value < 0.05 with darker shades indicating smaller *p*-values. e Principal component analysis of transcriptomic divergence (as AUC scores) between species for each cell type. AUC here quantifies the separability between species for each cell type and was calculated for each pair of species under control, ABA and NaCl treatments separately. Each symbol represents the transcriptomic divergence between six possible species pairs under one condition. The color and the shape of each symbol correspond to the cell type and the condition, respectively. f Transcriptomic distinctiveness (as AUC scores) between species pairs under different conditions. AUC here quantifies the separability between species for each cell type. AUC = 0.5 indicates random separability; AUC = 1.0 indicates perfect separability. Asterisks indicate significance after comparing to control (n = 50) with Student's t-test at *p*-value cutoff of 0.05. The red dash lines mark the medians the transcriptomic distinctiveness under control condition. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Sir, *Sisymbrium irio;* Spa, *Schrenkiella parvula*.

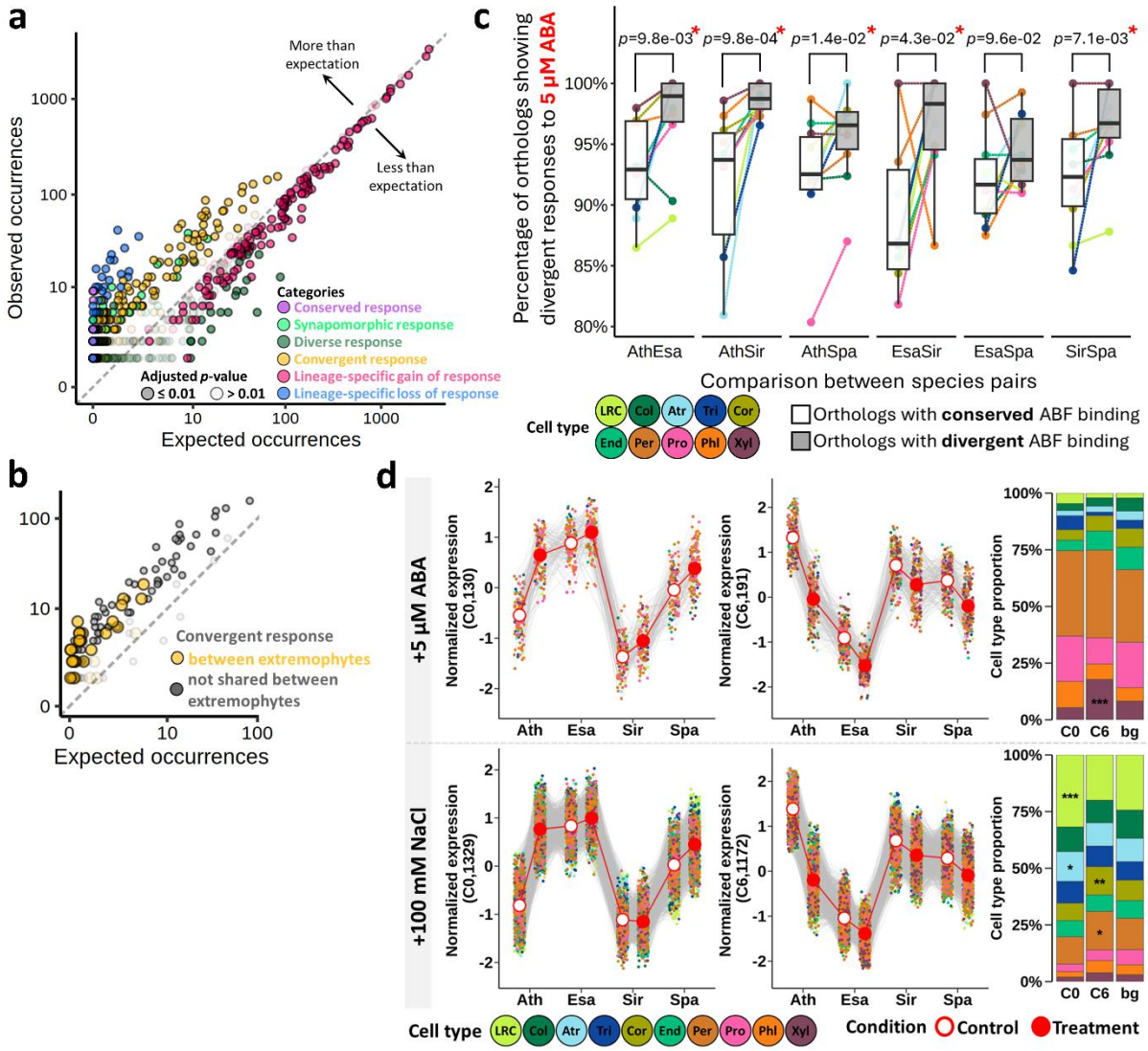


Fig. 6. Stress responses between orthologs. a-b Differences between observed and bootstrapped (10,000 times) occurrences of response patterns among all 1-to-1 orthologs (a) and orthologs showing convergent response (b). The darker and lighter color intensities represent significant and insignificant deviation of the observed occurrences from the expected distribution, respectively, at Benjamini-Hochberg-adjusted *p*-value ≤ 0.01. c Percentage of divergently regulated orthologs (in response to 5 μM ABA) with conserved or divergent ABF binding. ABF binding was defined as the presence of ABF binding events in the 2-kilobase upstream of the transcription start site using the DAP-seq data from Sun *et al.*, 2022[9]. See Supplementary Fig. 33a for more explanations. Numbers on the top represent *p*-values determined by the Wilcoxon test. Asterisks represent significance at the *p*-value cutoff of 0.05. d Contributions of each cell type to co-expression networks. Cluster ID and the number of orthologs in the cluster are shown in parentheses in the Y-axis labels. Each gray line represents the expression of an ortholog across conditions and species. The red line indicates the median of all orthologs in the given cluster. Cell type proportions of the cluster were compared to background (bg) where orthologs from all clusters were averaged (see Methods for details).

Asterisks represent significance determined by hypergeometric test, * *p*-value < 0.01, ** *p*-value < 0.001, *** *p*-value < 0.0001. LRC, Lateral Root Cap; Col, Columella; Atr, Atrichoblast; Tri, Trichoblast; Cor, Cortex; End, Endodermis; Per, Pericycle; Pro, Procambium; Phl, Phloem; Xyl, Xylem. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Sir, *Sisymbrium irio;* Spa, *Schrenkiella parvula*.

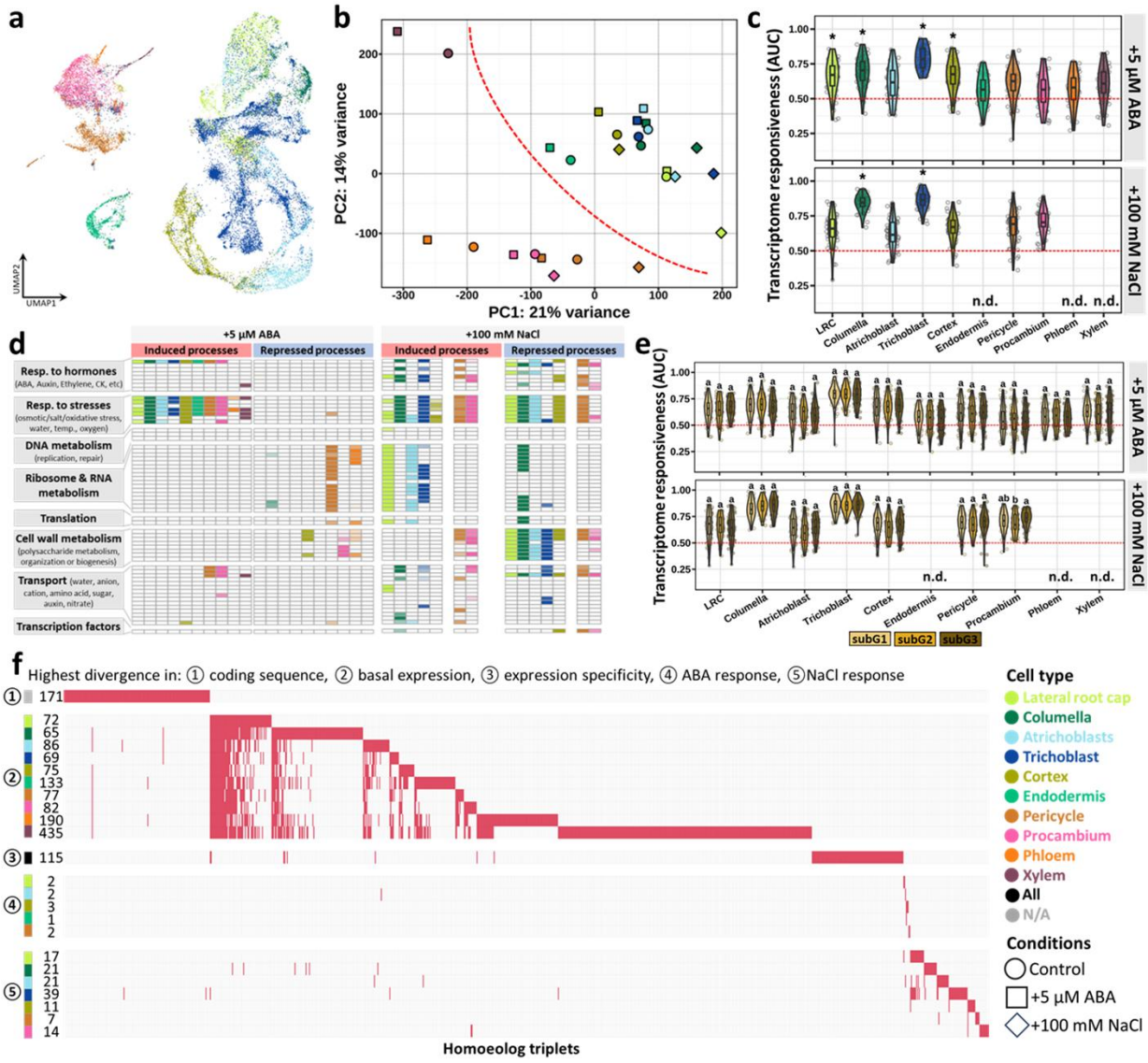
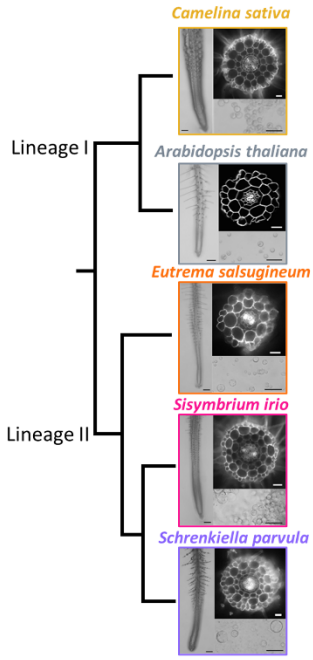


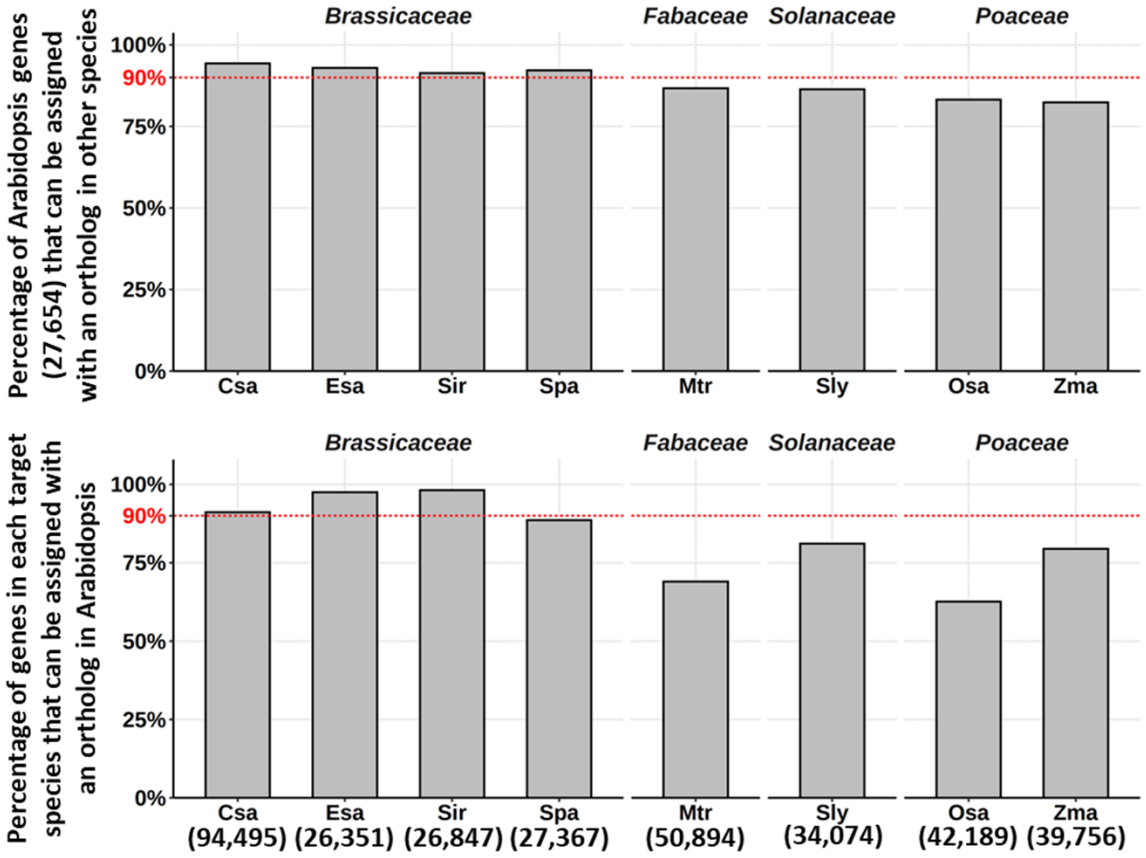
Fig. 7. Cell type stress responses in *C. sativa*. a Root single cell atlas (27,442 cells) for *C. sativa*. The atlas included control, 5 μM ABA- and 100 mM NaCl-treated samples. b Principal component analysis of pseudo-bulk cell-type transcriptomes. Each symbol represents a pseudo-bulk cell-type transcriptome from control, 5 μM ABA- and 100 mM NaCl-treated samples. The color and the shape of each symbol correspond to the cell type and the condition, respectively. c Comparison of transcriptome responsiveness as averaged Augur area under the curve (AUC) scores across cell types in response to ABA and NaCl. AUC here quantifies the separability between control and treatment samples for each cell type. AUC = 0.5 indicates random separability; AUC = 1.0 indicates perfect separability. Asterisks indicate significance after Wilcoxon testing (n = 50) at adjusted *p*-value cutoff of 0.05. d Summary of GO functional enrichment among DEGs combined from all sub-genomes in response to ABA (top) and NaCl (bottom) for each cell type in *C. sativa*. The detailed list of GO terms is present in Supplementary Data 7. The shade represents the *p*-value assigned by the enrichment test at

adjusted *p*-value < 0.05 with darker shades indicating smaller *p*-values. e Comparison of transcriptome responsiveness (as AUC scores) for each sub-genome in *C. sativa*. AUC here quantifies the separability between control and treatment samples using each sub-genome for each cell type. AUC = 0.5 indicates random separability; AUC = 1.0 indicates perfect separability. Lowercase letters on top of violin plots indicate significantly different groups (n = 50) as determined by one-way analysis of variance (ANOVA) with post-hoc Tukey honest significant difference (HSD) testing at *p*-value < 0.01. f Ranked homoeologous triplets based on their divergence in coding sequence (①), expression (②), cell type specificity (③) as well as expression changes in response to stresses (④ and ⑤). Numbers on the left indicate the number of candidates considered for each category. Endodermis, phloem and xylem cells were removed from the comparison for NaCl treatment (n.d.) due to insufficient number of cells (< 50 cells) in c-f. N/A, not applicable.

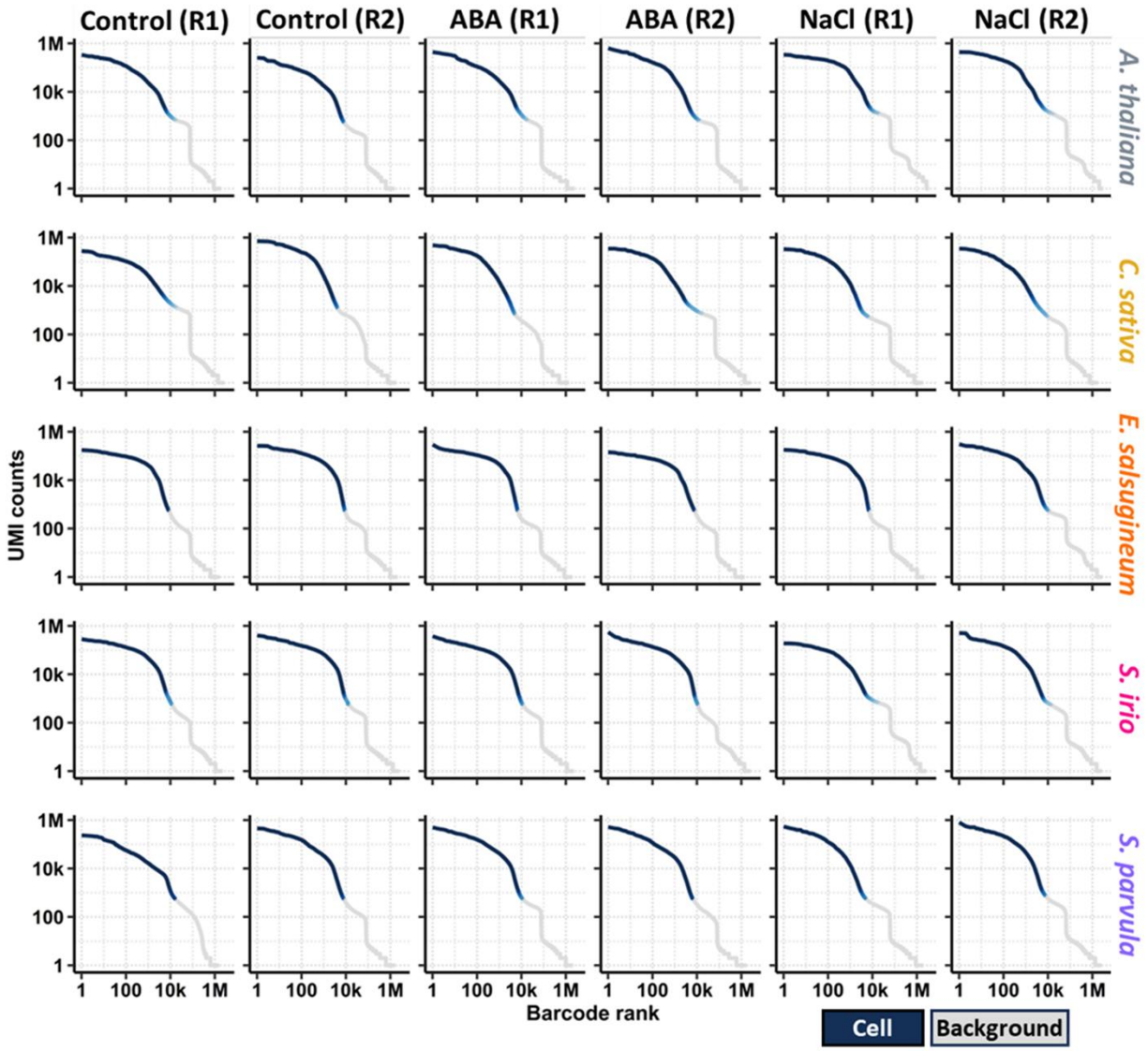


Input cell counts per conditions (2 replicates)			Preprocessing	Filtered cell counts per condition (2 replicates)			Atlas
Control	ABA	NaCl		Control	ABA	NaCl	
3,848 + 8,871	4,109 + 6,583	3,974 + 4,538	➢ Comparison between pseudo-bulk RNA-seq and whole root RNA-seq. ➢ Evaluation of the effects of protoplasting. ➢ Removal of cells with high mitochondrial counts. ➢ Removal of doublets. ➢ Check for consistency between replicates.	2,947 + 7,727	3,489 + 5,943	3,365 + 3,971	27,442
7,551 + 9,963	9,885 + 8,795	10,615 + 7,627		6,604 + 8,817	8,419 + 7,880	8,953 + 6,355	47,028
8,633 + 8,608	6,498 + 7,869	6,788 + 8,477		7,733 + 7,648	5,775 + 6,843	5,975 + 7,566	41,540
10,719 + 8,992	10,039 + 9,316	8,155 + 7,894		9,800 + 8,099	9,077 + 8,455	7,584 + 7,219	50,234
7,755 + 16,980	10,010 + 6,619	4,435 + 6,986		6,282 + 6,345	8,255 + 5,497	3,649 + 5,990	36,018

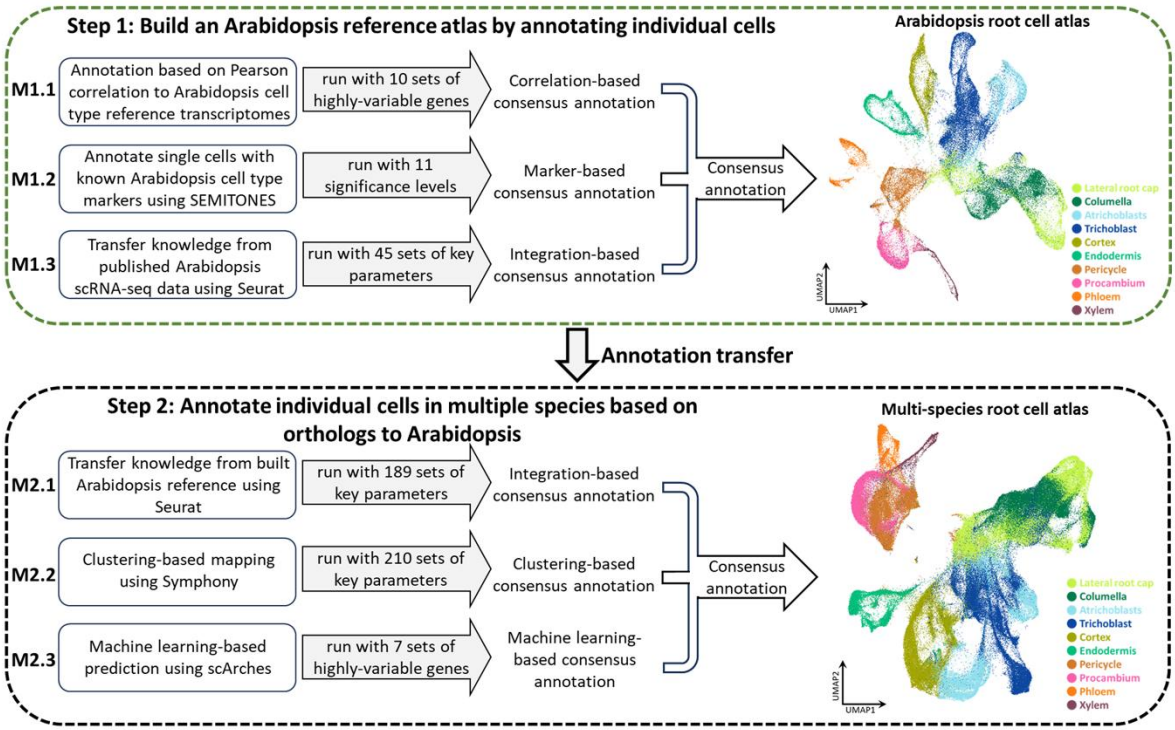
Supplementary Fig. 1. Overview of scRNA-seq dataset for four 5 Brassicaceae species and data preprocessing. The number in each UMAP represents the total number of filtered cells from control, 5 µM ABA- and 100 mM NaCl-treated samples.



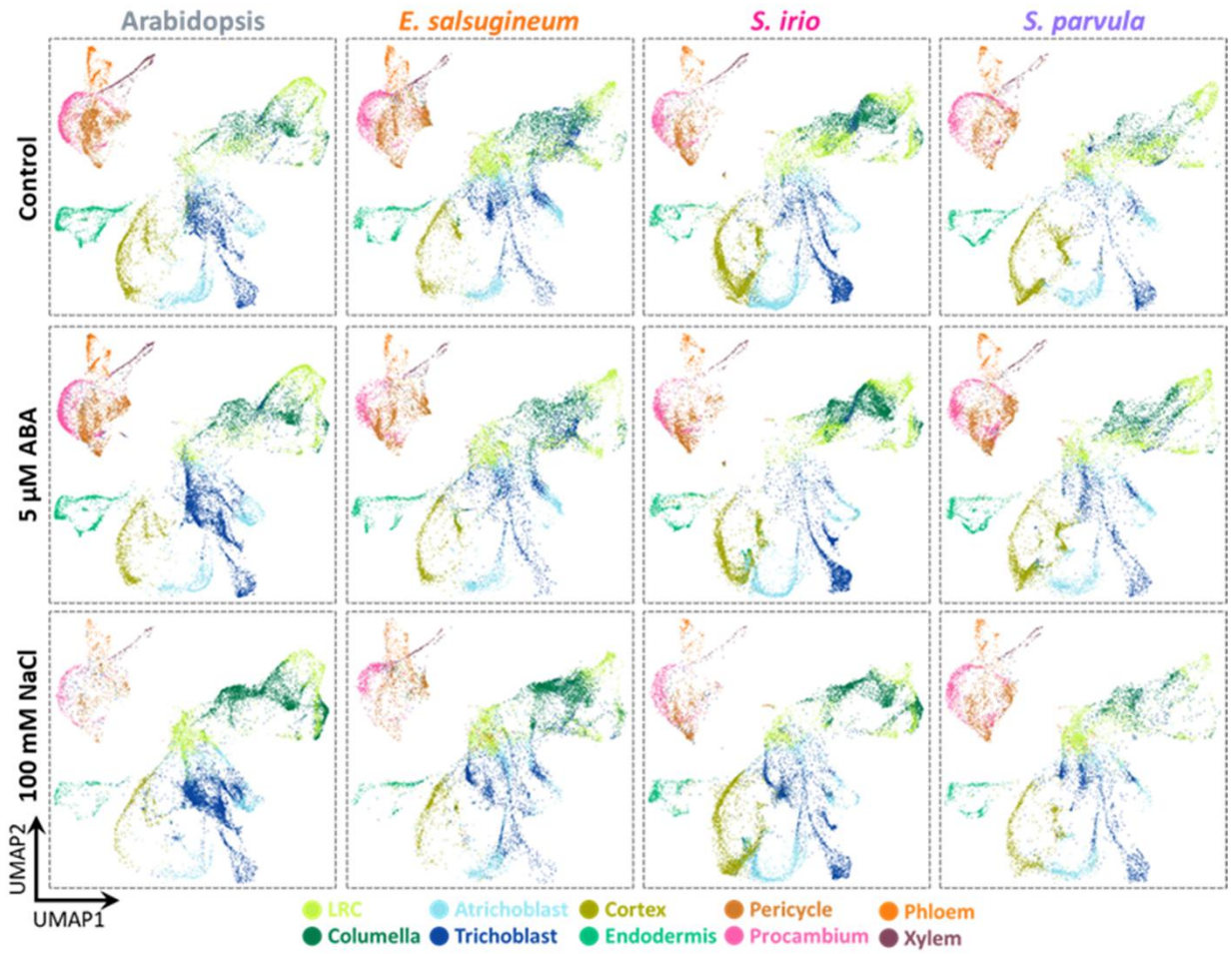
Supplementary Fig. 2. Percentage of genes with orthologs between Arabidopsis and other species. Csa, *Camelina sativa*; Esa, *Eutrema salsugineum*; Spa, *Schrenkiella parvula*; Sir, *Sisymbrium irio*; Mtr, *Medicago truncatula*; Sly, *Solanum lycopersicum*; Osa, *Oryza sativa*; Zma, *Zea mays*. Numbers in parenthesis indicate the numbers of protein-coding genes for each species.



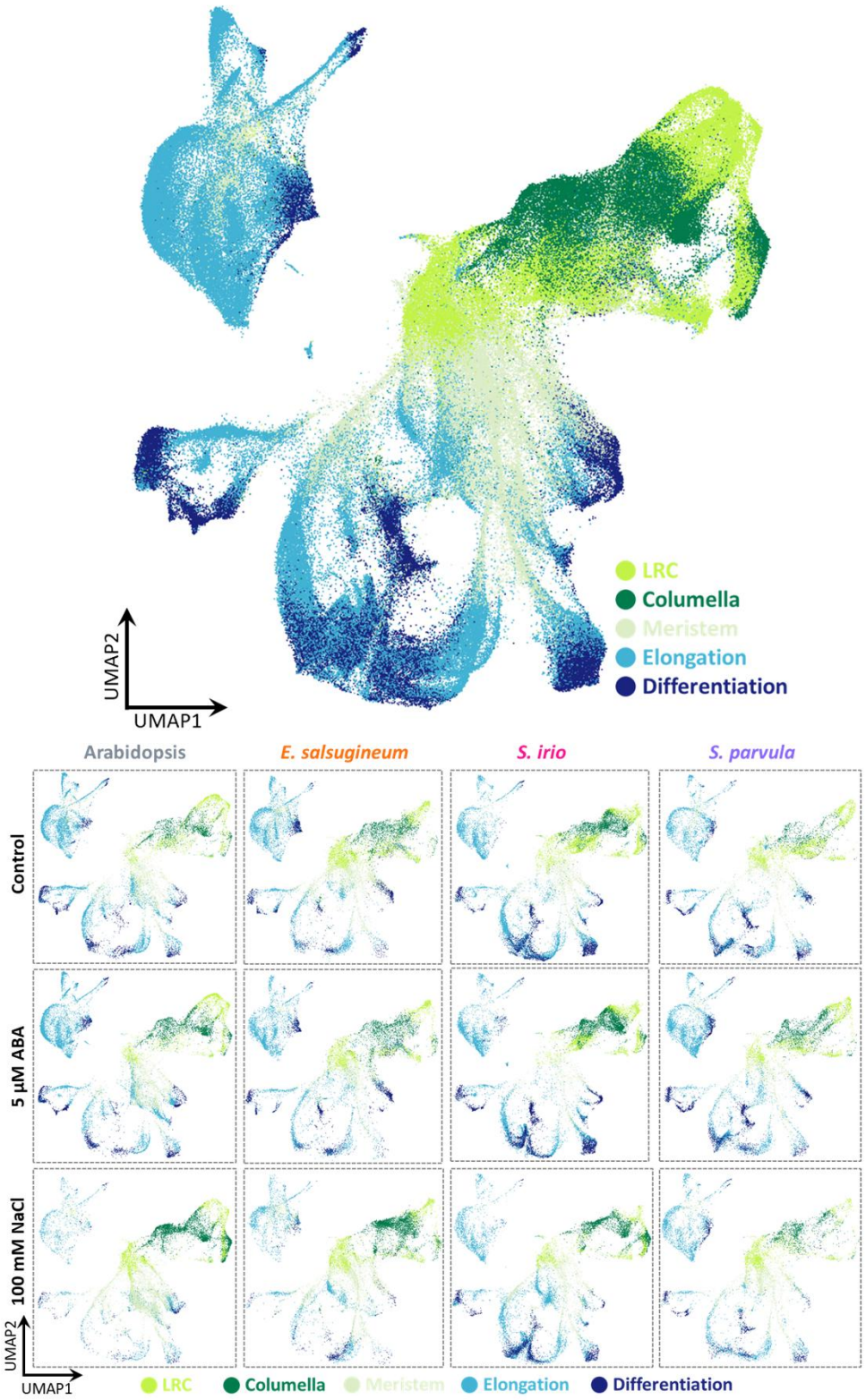
Supplementary Fig. 3. Knee plots for each sample in our study. For each species, samples were harvested from control, 5μM ABA-, and 100 mM NaCl-treated conditions. Two independent replicates (R1 and R2) were used for each condition. The high-confidence cell-associated barcodes were clearly separated from the background ones.



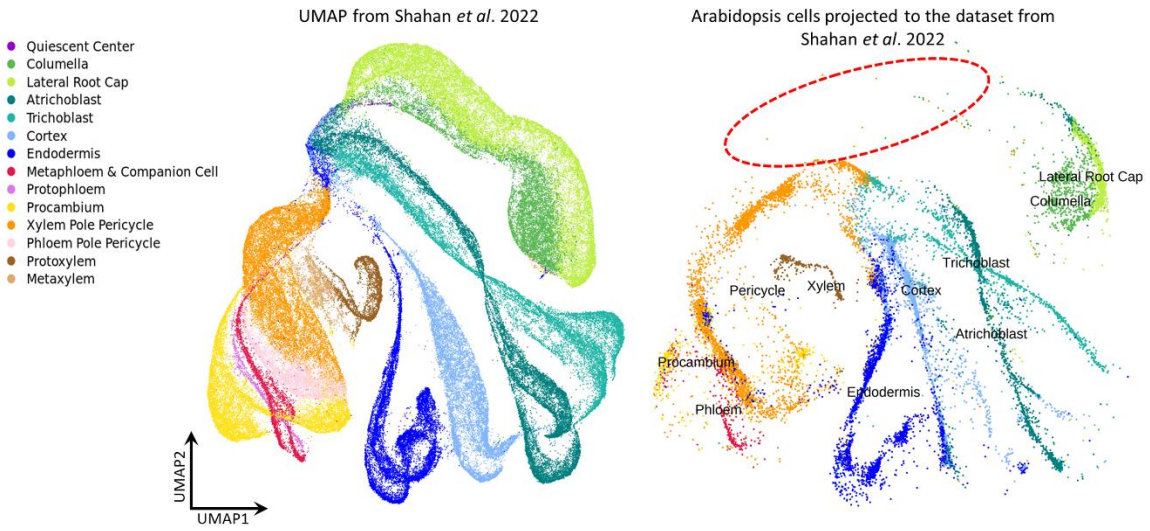
Supplementary Fig. 4. A two-step strategy for cross-species cell type annotations using three independent methods. The Arabidopsis cell type reference transcriptomes were from Li *et al.*, 2016[64] and Brady *et al.*, 2007[65]. The different sets of highly-variable genes were obtained by categorizing them into different percentile bins. The key parameters for each method were selected based on the tools used, see details in Methods.



Supplementary Fig. 5. Root cell atlas for each species under different conditions on the integrated UMAP space. All ten root cell types were detected in each species under different conditions. LRC, Lateral Root Cap.

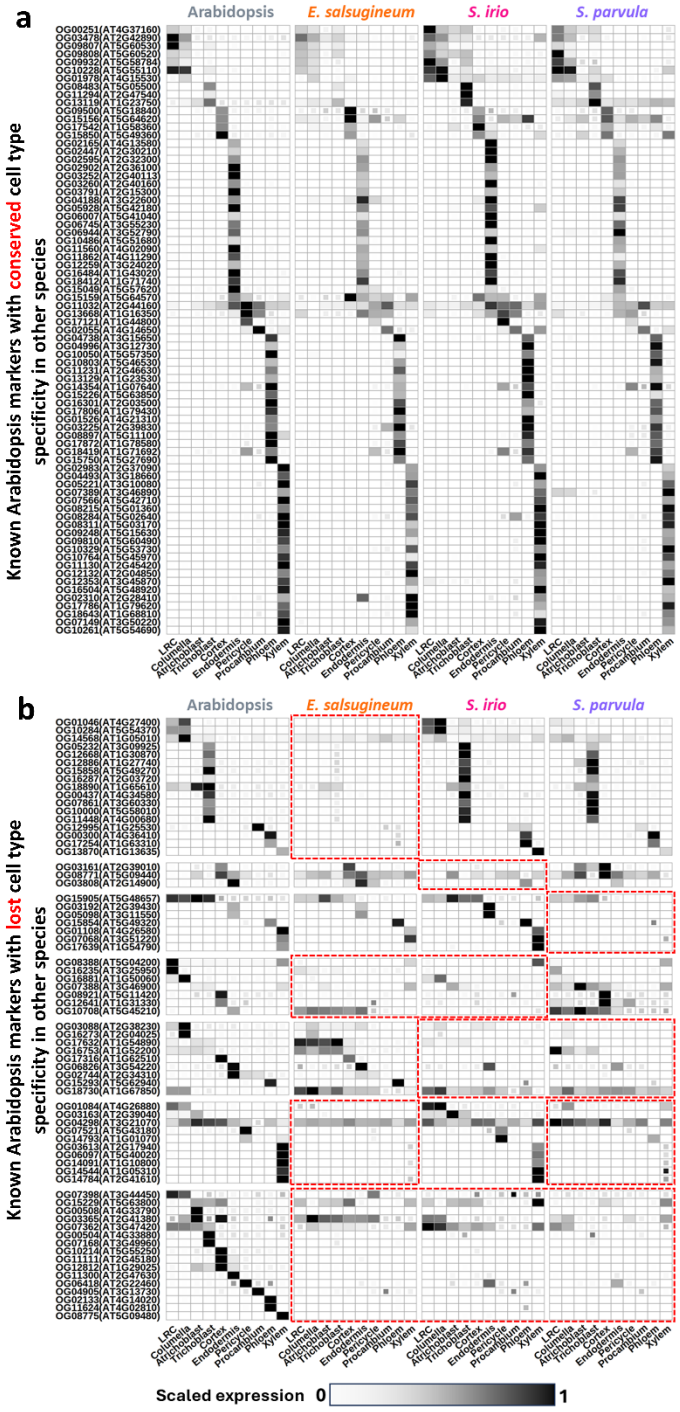


Supplementary Fig. 6. Developmental zone annotation for the integrated root cell atlas from four species (top) and in each species (bottom). All developmental zones were detected in each species under different conditions. LRC, Lateral Root Cap.

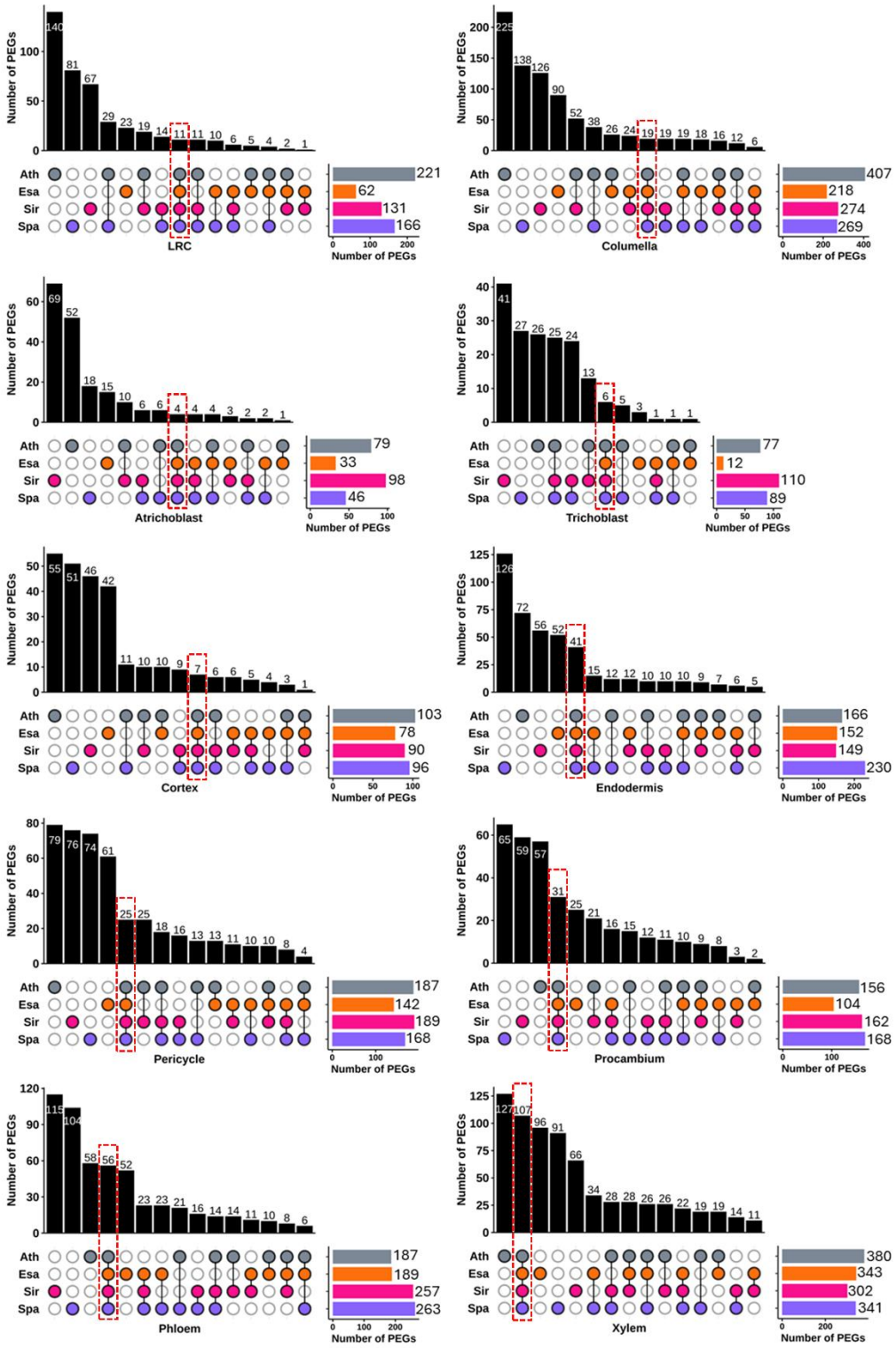


Supplementary Fig. 7. Projection of Arabidopsis control cells from current study to Shahan *et al*[21]. The red dashed oval marks the expected space for quiescent center cells, currently not detected in the current dataset for Arabidopsis.

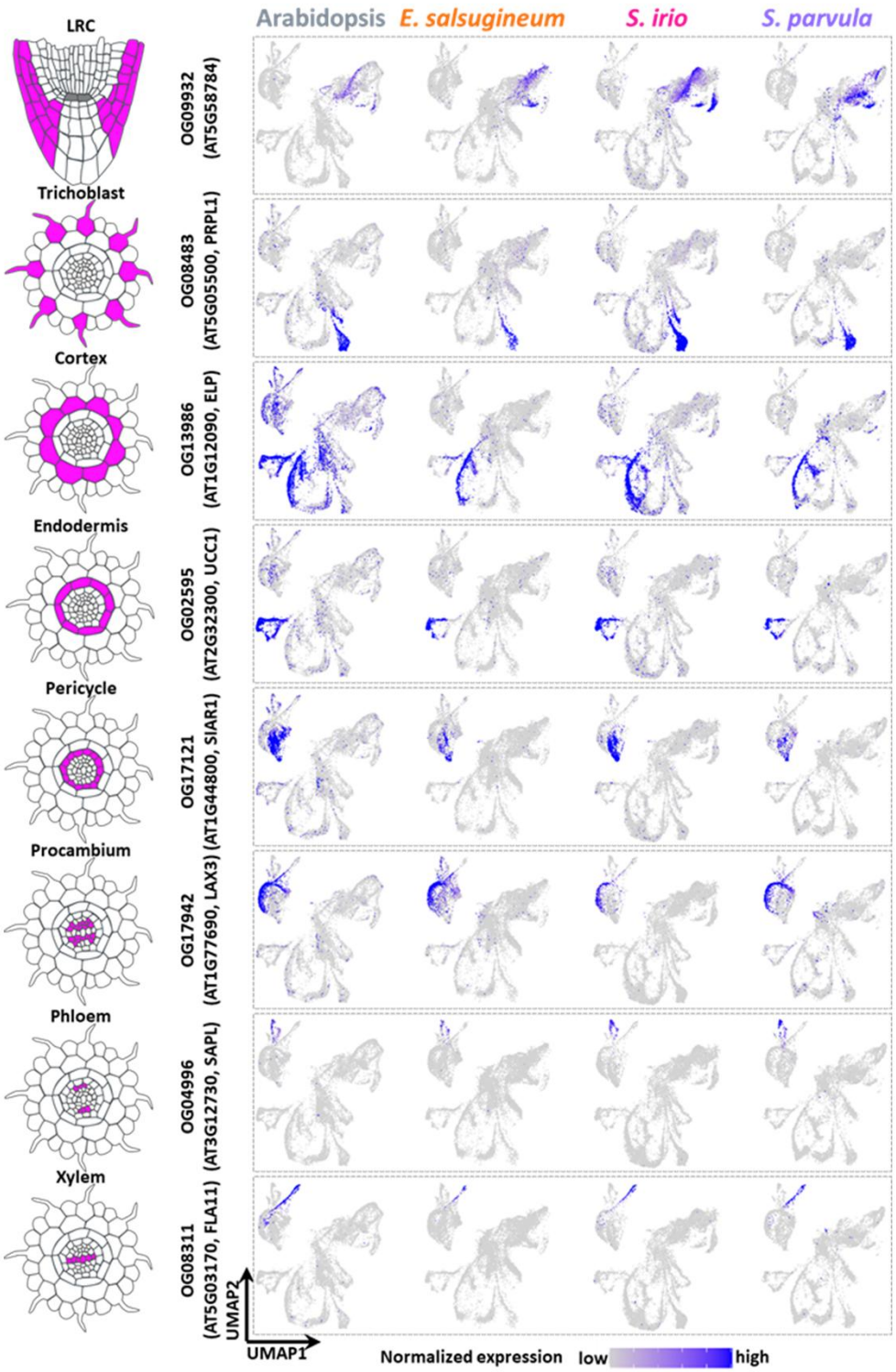
Supplementary Fig. 8. Cell-type expression of known Arabidopsis cell-type markers and their 1-to-1 ortholog in other species under control condition. See Supplementary Data 2 for the list of markers used. Dot size represents the percentage of cells in which each gene is expressed. Dot color intensity indicates the normalized expression of each gene in each cell type. QC, Quiescent center; LRC, Lateral Root Cap; Col, Columella; Atr, Atrichoblast; Tri, Trichoblast; Cor, Cortex; End, Endodermis; Per, Pericycle; Pro, Procambium; Phl, Phloem; Xyl, Xylem.



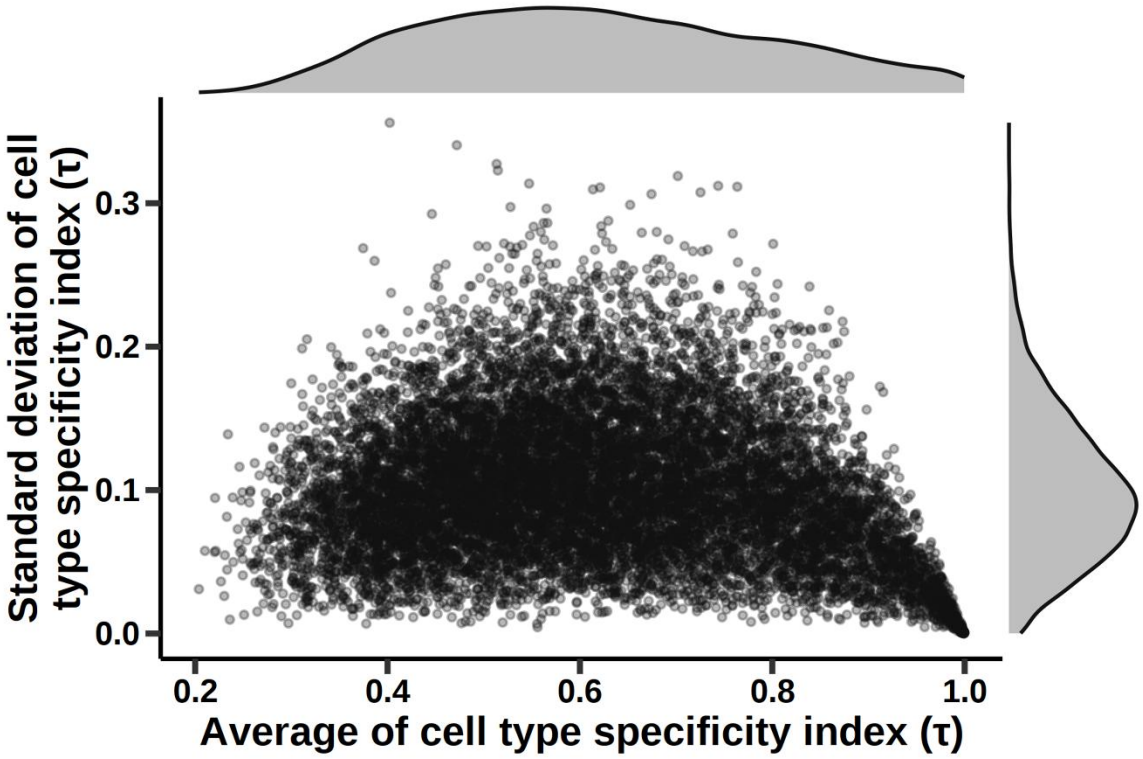
Supplementary Fig. 9. Expression of conserved (a) and diverged (b) Arabidopsis cell type markers across species under control condition. Red boxes mark the species where these cell type markers lost their expression specificity. Tiles with incomplete fill indicate these genes were expressed in less than 10% of cells from the target cell types. They were not considered for marker identification.



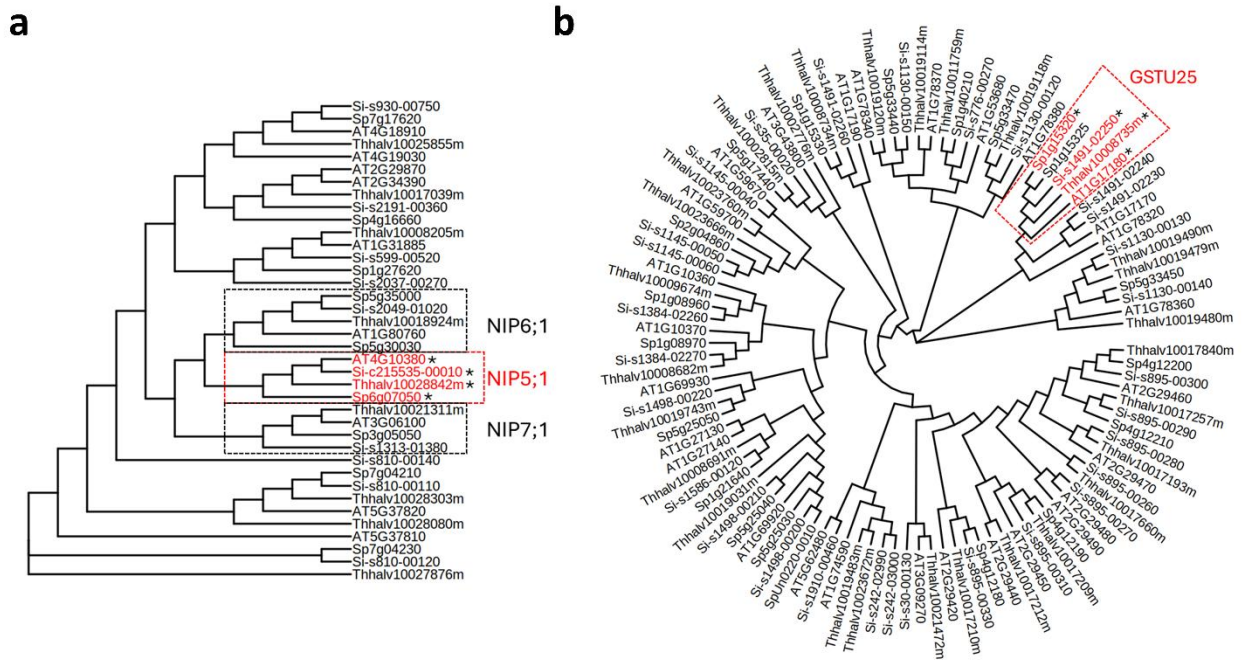
Supplementary Fig. 10. Identification of cell type preferentially enriched genes (PEGs) that were conserved across species for each cell type in the current study. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Spa, *Schrenkiella parvula*; Sir, *Sisymbrium irio*. The conserved PEGs for each cell type are highlighted with red boxes.

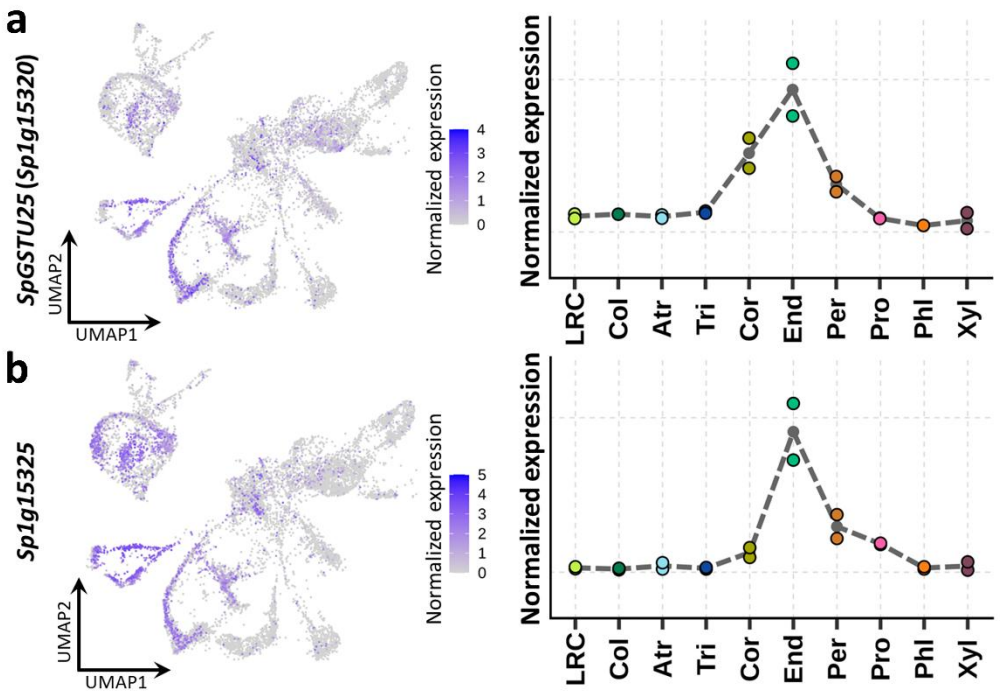


Supplementary Fig. 11. Expression of selected cell type markers in Fig. 2 across species. The expression shown is from the control samples. PRPL1, proline-rich protein-like 1; ELP, extensin-like protein; UCC1, uclacyanin 1; SIAR1, siliques are red 1; LAX3, like aux1 3; SAPL, Sister of APL; FLA11, FASCICLIN-like arabinogalactan-protein 11.

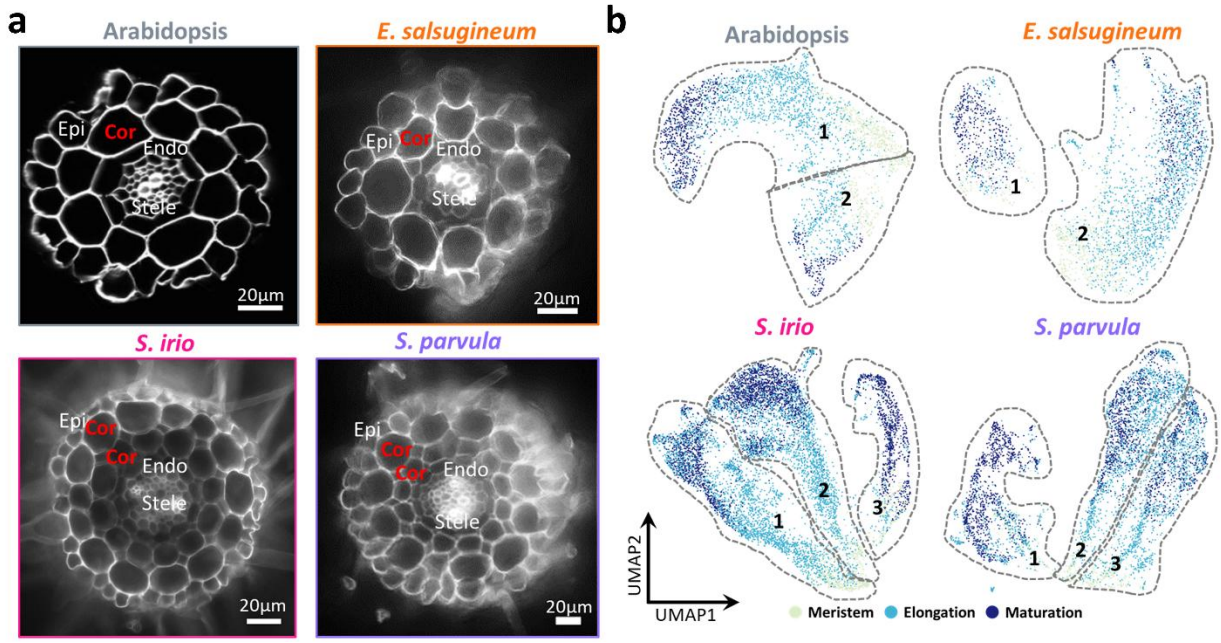


Supplementary Fig. 12. Comparison of the average and the standard deviation of cell type specificity index for the 1-to-1 orthologs. Cell type specificity index (τ) is based on tau metric of tissue specificity in the control samples of each species and calculated as $\sum_{i=1}^{N}(1-x_i)/(N-1)$, where N is the number of cell types and x_i is the expression level in cell type is normalized by the maximal expression of the given gene across all cell types. Cell type specificity index (τ) ranges from 0 to 1, where 0 means broad expression in all cell types and 1 indicates specific expression in one cell type. The cell type specificity indexes for each 1-to-1 orthologs were used to calculate the average and standard deviation for these orthologs.

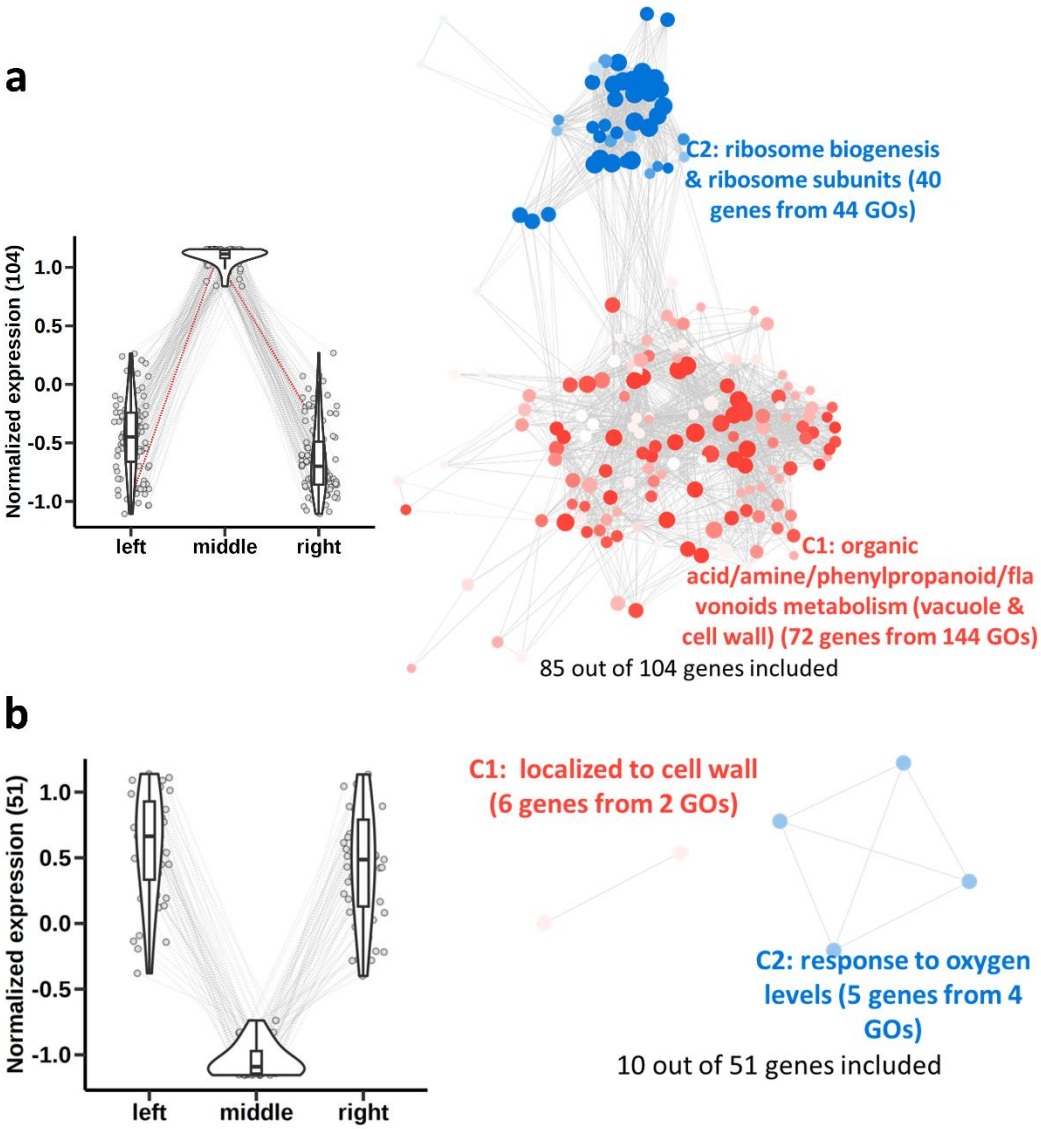

a
Si-s930-00750
Sp7g17620
AT4G18910
Thhalv10025855m
AT4G19030
AT2G29870
AT2G34390
Thhalv10017039m
Si-s2191-00360
Sp4g16660
Thhalv10008205m
AT1G31885
Si-s599-00520
Sp1g27620
Si-s2037-00270
Sp5g35000
Si-s2049-01020
Thhalv10018924m
AT1G80760
Sp5g30030
AT4G10380
Si-c215535-00010
Thhalv10028842m
Sp6g07050
Thhalv10021311m
AT3G06100
Sp3g05050
Si-s1313-01380
Si-s810-00140
Sp7g04210
Si-s810-00110
Thhalv10028303m
AT5G37820
Thhalv10028080m
AT5G37810
Sp7g04230
Si-s810-00120
Thhalv10027876m
NIP6;1
NIP5;1
NIP7;1
b
GSTU25
Si-s1491-02250
Thhalv10008735m
AT1G17180
Si-s1491-02240
Si-s1491-02230
Si-s1491-02260
Si-s1130-00120
Si-s1130-00130
Si-s1130-00140
Si-s1130-00150
Si-s776-00270
Si-s895-00300
Si-s895-00290
Si-s895-00280
Si-s895-00260
Si-s895-00270
Si-s895-00310
Si-s895-00330
Si-s895-00210
Si-s1498-00220
Si-s1498-00210
Si-s1498-00200
Si-s1910-00460
Si-s242-02990
Si-s242-03000
Si-s30-00130
Si-s1384-02260
Si-s1384-02270
Si-s1145-00050
Si-s1145-00060
Si-s1145-00040
Si-s35-00020
Si-s1499-00220
Si-s1586-00120
Si-s1313-01380
Si-s2049-01020
Si-s599-00520
Si-s2191-00360
Si-s930-00750
Si-s810-00140
Si-s810-00110
Si-s810-00120
Si-s2037-00270
Si-s776-00270
Si-s1498-00200
Si-s1498-00210
Si-s1498-00220
Si-s1910-00460
Si-s242-02990
Si-s242-03000
Si-s30-00130
Si-s895-00300
Si-s895-00290
Si-s895-00280
Si-s895-00260
Si-s895-00270
Si-s895-00310
Si-s895-00330
Si-s895-00210
Si-s1491-02250
Si-s1491-02240
Si-s1491-02230
Si-s1491-02260
Si-s1130-00120
Si-s1130-00130
Si-s1130-00140
Si-s1130-00150
Si-s776-00270
Si-s895-00300
Si-s895-00290
Si-s895-00280
Si-s895-00260
Si-s895-00270
Si-s895-00310
Si-s895-00330
Si-s895-00210
Si-s1498-00220
Si-s1498-00210
Si-s1498-00200
Si-s1910-00460
Si-s242-02990
Si-s242-03000
Si-s30-00130
Si-s1384-02260
Si-s1384-02270
Si-s1145-00050
Si-s1145-00060
Si-s1145-00040
Si-s35-00020
Si-s1499-00220
Si-s1586-00120
Si-s1313-01380
Si-s2049-01020
Si-s599-00520
Si-s2191-00360
Si-s930-00750
Si-s810-00140
Si-s810-00110
Si-s810-00120
Si-s2037-00270
Si-s776-00270
Si-s1498-00200
Si-s1498-00210
Si-s1498-00220
Si-s1910-00460
Si-s242-02990
Si-s242-03000
Si-s30-00130
Si-s895-00300
Si-s895-00290
Si-s895-00280
Si-s895-00260
Si-s895-00270
Si-s895-00310
Si-s895-00330
Si-s895-00210
Si-s1491-02250
Si-s1491-02240
Si-s1491-02230
Si-s1491-02260
Si-s1130-00120
Si-s1130-00130
Si-s1130-00140
Si-s1130-00150
Si-s776-00270
Si-s895-00300
Si-s895-00290
Si-s895-00280
Si-s895-00260
Si-s895-00270
Si-s895-00310
Si-s895-00330
Si-s895-00210
Si-s1498-00220
Si-s1498-00210
Si-s1498-00200
Si-s1910-00460
Si-s242-02990
Si-s242-03000
Si-s30-00130
Si-s1384-02260
Si-s1384-02270
Si-s1145-00050
Si-s1145-00060
Si-s1145-00040
Si-s35-00020
Si-s1499-00220
Si-s1586-00120
Si-s1313-01380
Si-s2049-01020
Si-s599-00520
Si-s2191-00360
Si-s930-00750
Si-s810-00140
Si-s810-00110
Si-s810-00120
Si-s2037-00270
Si-s776-00270
Si-s1498-00200
Si-s1498-00210
Si-s1498-00220
Si-s1910-00460
Si-s242-02990
Si-s242-03000
Si-s30-00130
Thhalv10017840m
Sp4g12200
Thhalv10017257m
Sp4g12210
Thhalv10017193m
Thhalv10017660m
Thhalv10017209m
Thhalv10017212m
Thhalv10017210m
Thhalv10021472m
Thhalv10019480m
Thhalv10019479m
Thhalv10019490m
Thhalv10019120m
Thhalv10019114m
Thhalv10011759m
Thhalv10019118m
Thhalv10008682m
Thhalv10019743m
Thhalv10009674m
Thhalv10023666m
Thhalv10023760m
Thhalv10027876m
Thhalv10028842m
Thhalv10018924m
Thhalv10028080m
Thhalv10028303m
Thhalv10025855m
Thhalv10017039m
Thhalv10008205m
Thhalv10021311m
Thhalv10008691m
Thhalv10019031m
Thhalv10023672m
Thhalv10019479m
Thhalv10019490m
Thhalv10019480m
Thhalv10017840m
Thhalv10017257m
Thhalv10017193m
Thhalv10017660m
Thhalv10017209m
Thhalv10017212m
Thhalv10017210m
Thhalv10021472m
Thhalv10027876m
Thhalv10028303m
Thhalv10028080



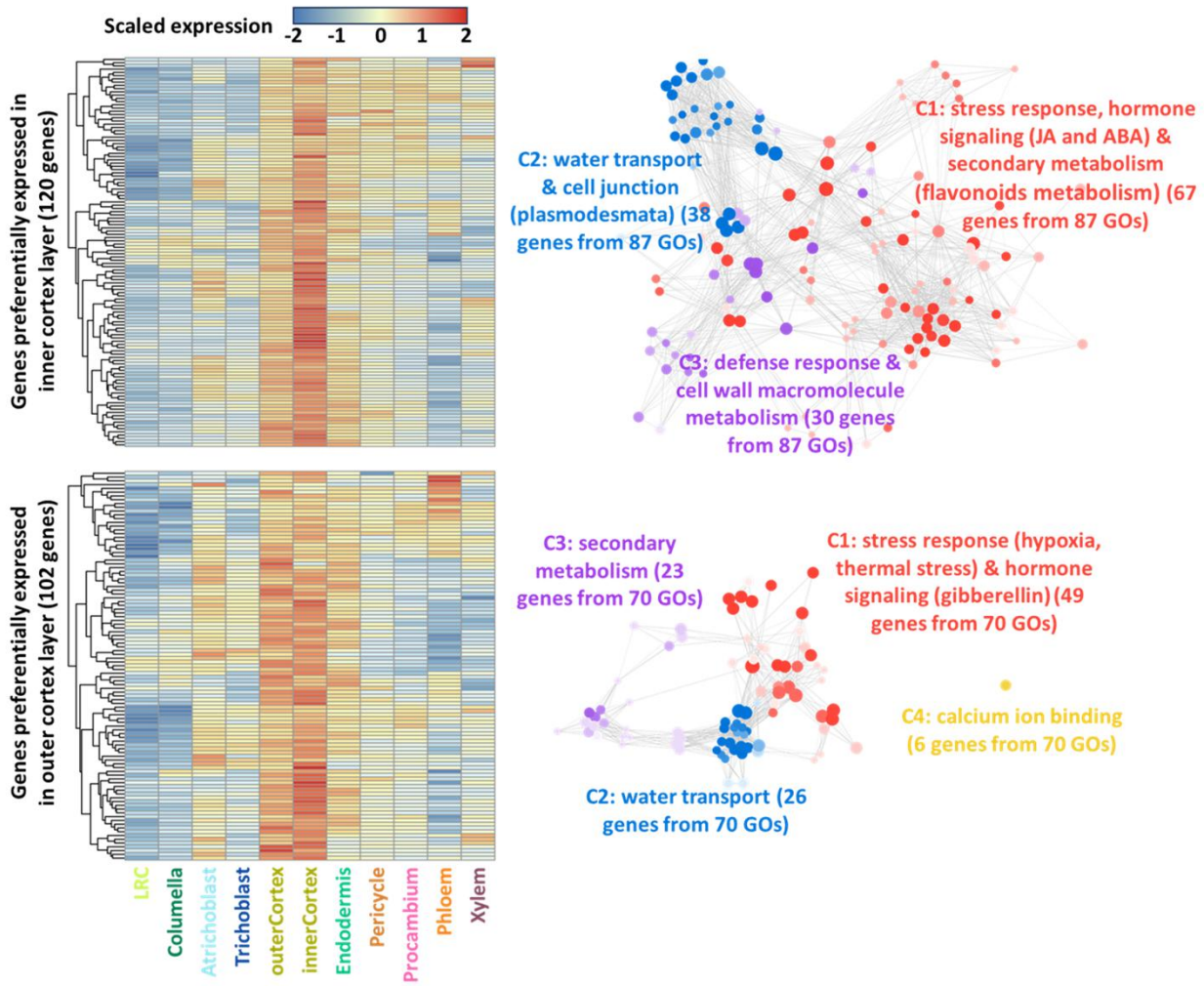
Supplementary Fig. 14. Expression of *SpGSTU25* (*Sp1g15320*) (a) and *Sp1g15325* (b) based on scRNA-seq data for *S. parvula* under control condition. Left panels display the expression of the selected genes visualized on UMAP space. Right panels show the pseudo-bulk cell type expression of the selected genes. LRC, Lateral Root Cap; Col, Columella; Atr, Atrichoblast; Tri, Trichoblast; Cor, Cortex; End, Endodermis; Per, Pericycle; Pro, Procambium; Phl, Phloem; Xyl, Xylem.



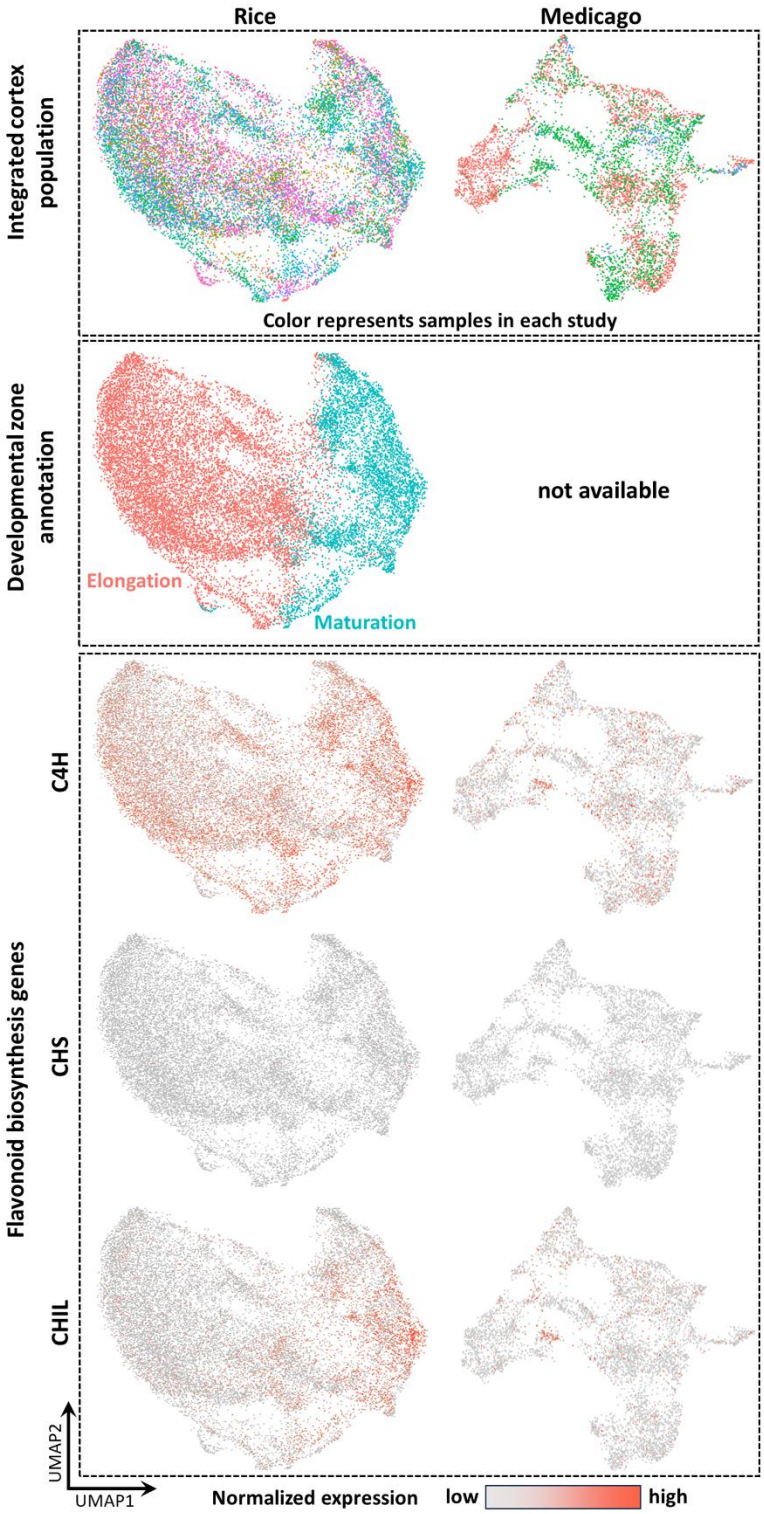
Supplementary Fig. 15. Comparison of cortex cells across species from control samples. a Cross-sections of roots at the maturation zone in each species. Roots were obtained from 5–6-day old seedlings for each species. Scale bars, 20 µm. b UMAP representation of cortex cells in each species. Sub-populations in each species are manually circled based on trajectories identified by Monocle.



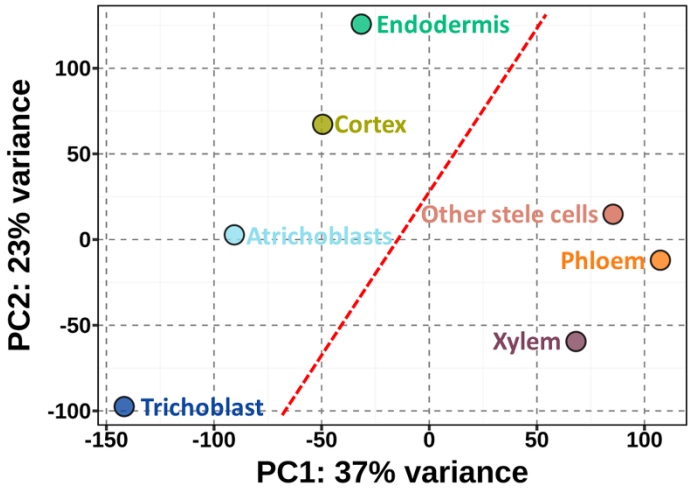
Supplementary Fig. 16. Functional representations of genes that were preferentially expressed (a) and depleted (b) in the middle cortex subpopulation of *S. irio*. Numbers in the parenthesis on the y-axis indicate the number of genes. Box plots represent the median (central value), upper and lower quartile (box limits) with whiskers at ± 1.5 × the interquartile range. The red dash line in the line graph in (a) marks the normalized expression of *Si-s1384-01650*, the HCR result of its expression in *S. irio* is shown in Fig. 4d. GO clusters are differently colored and labeled with the representative functional term. Each node represents a gene ontology (GO) term; node size represents genes in the test set assigned to that functional term; GO terms sharing > 50% of genes are connected with edges; and the shade of each node represents the *p*-value assigned by the enrichment test at adjusted *p* values < 0.05 with darker shades indicating smaller *p*-values.



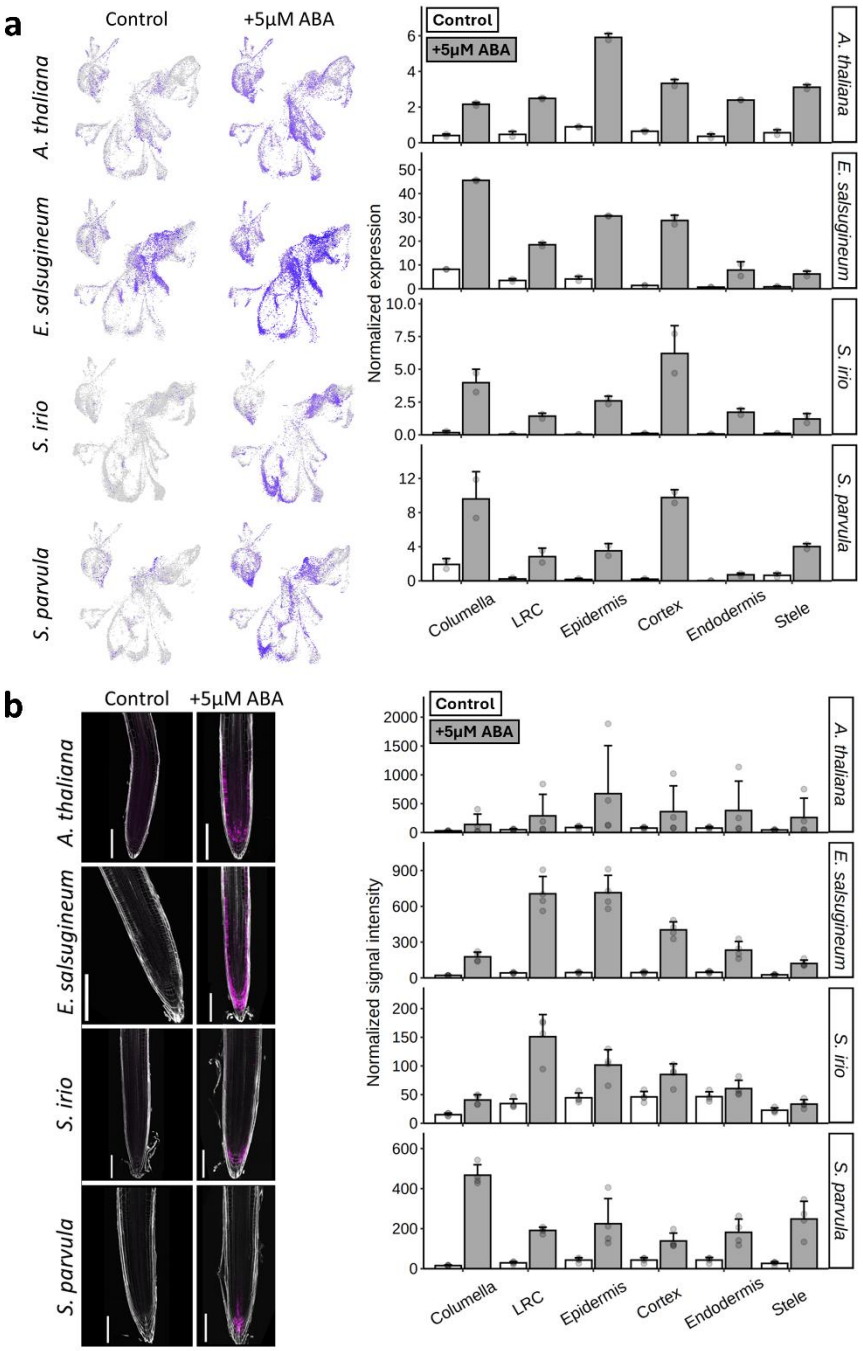
Supplementary Fig. 17. Expression of genes that are preferentially expressed in inner (top) and outer (bottom) cortex layers in *S. irio* and their functional representations. GO clusters are differently colored and labeled with the representative functional term. Each node represents a gene ontology (GO) term; node size represents genes in the test set assigned to that functional term; GO terms sharing > 50% of genes are connected with edges; and the shade of each node represents the p-value assigned by the enrichment test at adjusted p values < 0.05 with darker shades indicating smaller p-values.



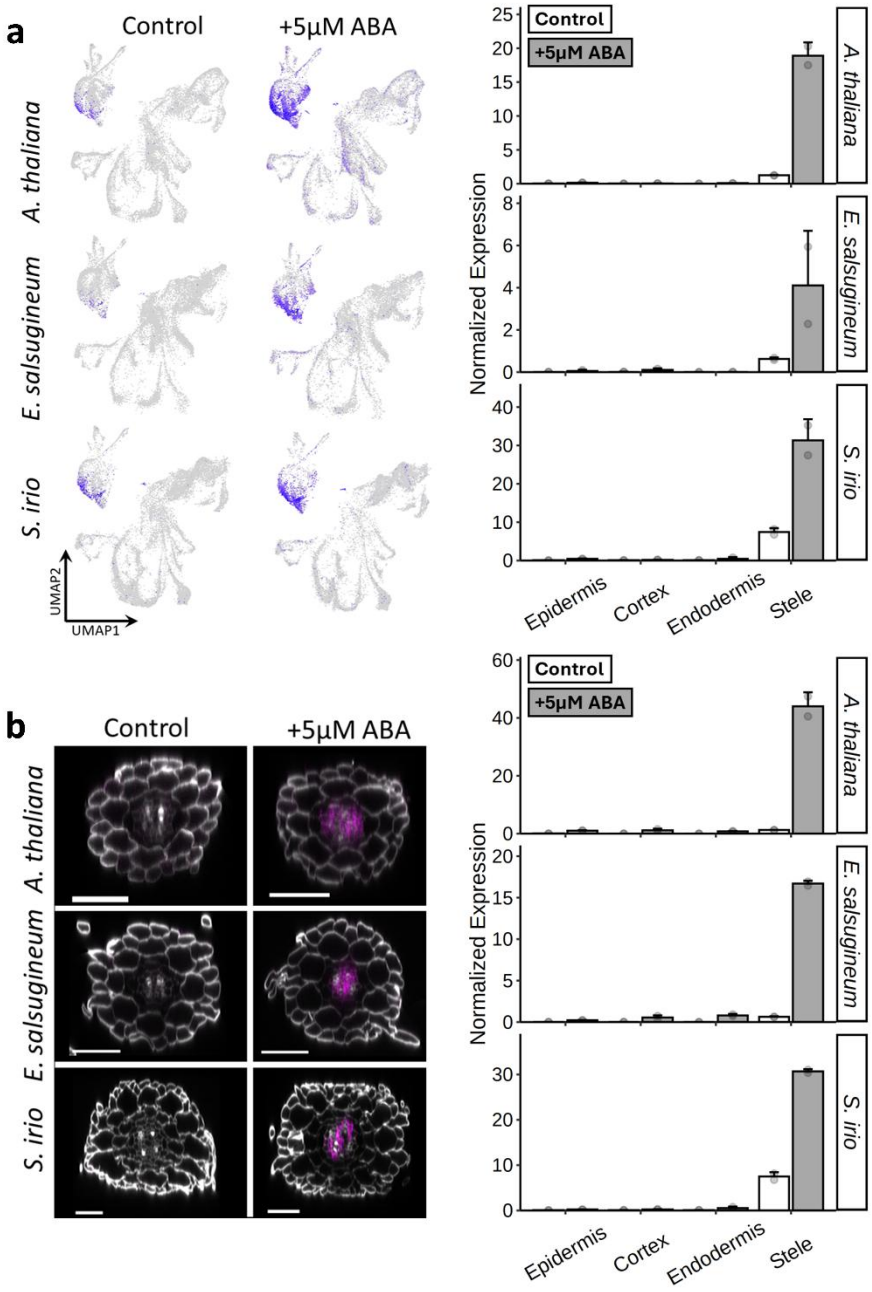
Supplementary Fig. 18. The expression of flavonoid biosynthetic genes in rice and Medicago. The spatial separation of the flavonoid pathway in different cortex layers is not conserved in rice and Medicago. C4H, cinnamic acid 4-hydroxylase; CHS, chalcone synthase; CHIL, chalcone isomerase-like.



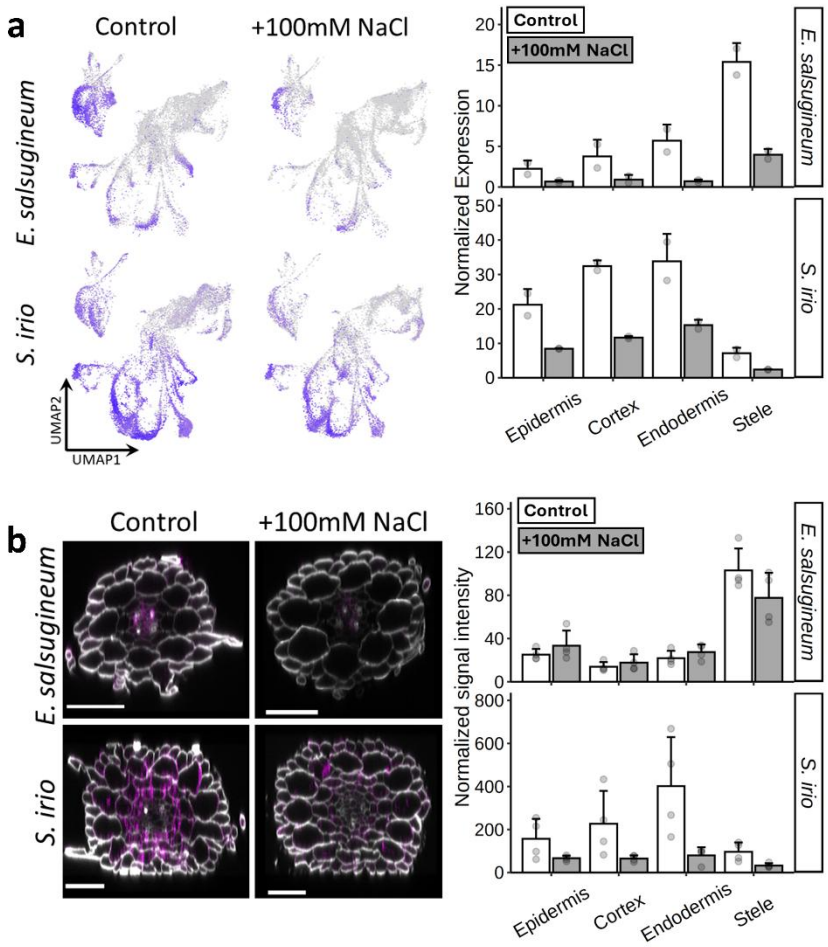
Supplementary Fig. 19. Principal component (PC) analysis of pseudo-bulk cell-type transcriptomes generated from single-nucleus RNA-seq of Arabidopsis (Farmer *et al*., 2021[37]). Each symbol represents a pseudo-bulk cell-type transcriptome. The red dashed lines separate cell type transcriptome subgroups within each species.



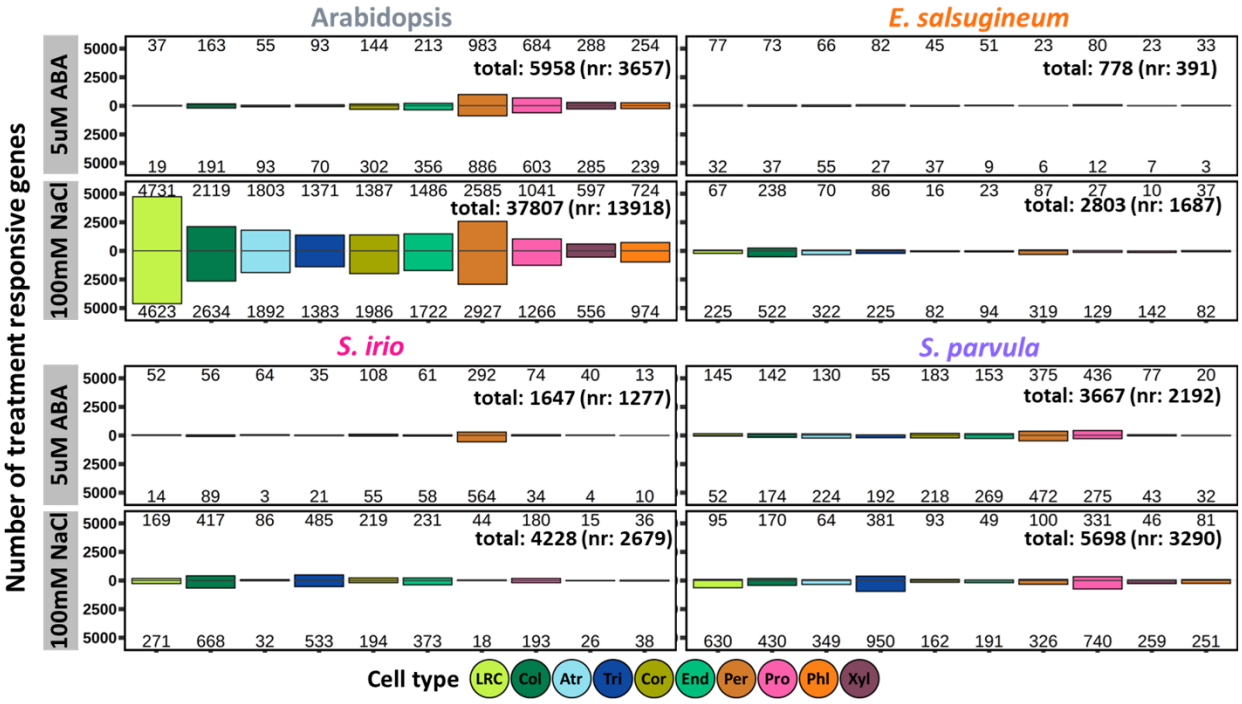
Supplementary Fig. 20. Expression changes of OG04736 orthologs (*AT3G15670*) across species in response to ABA. a Expression of OG04736 orthologs across species from scRNA-seq data under different conditions. Left, UMAP representation of the expression of OG04736 orthologs. Right, pseudo-bulk cell type expression of OG04736 orthologs. b Whole-mount *in situ* HCR detection of expression changes of OG04736 orthologs across species in response to ABA. Left, confocal images of whole-mount HCR showing the detection of OG04736 orthologs across species. Scale bars, 150 μm. Right, average signal intensity of OG04736 orthologs across defined cell type groups.



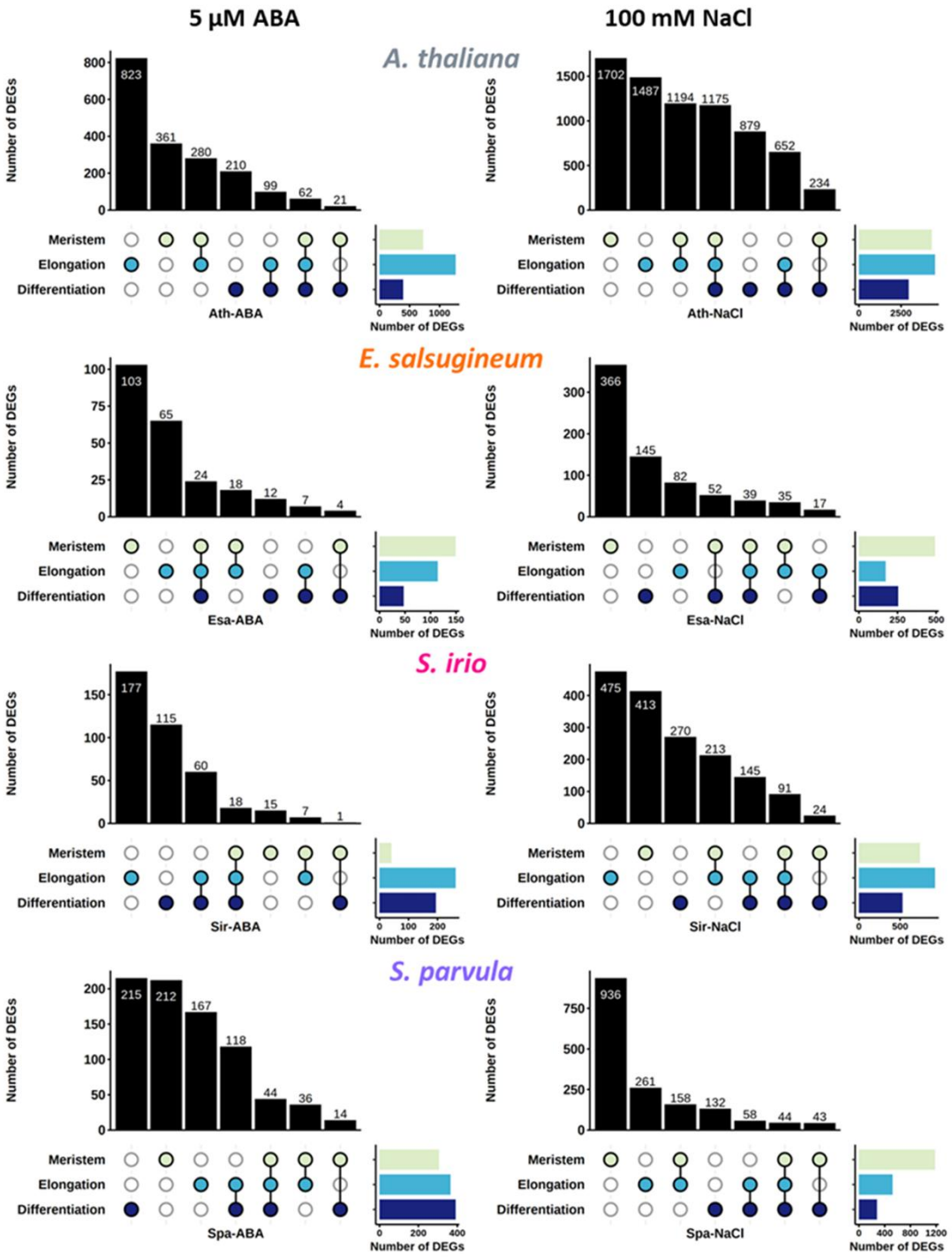
Supplementary Fig. 21. Expression changes of OG03053 orthologs (*AT2G37870*) across species in response to ABA. a Expression of OG03053 orthologs across species from scRNA-seq data under different conditions. Left, UMAP representation of the expression of OG03053 orthologs. Right, pseudo-bulk cell type expression of OG03053 orthologs. b Whole-mount *in situ* HCR detection of expression changes of OG03053 orthologs across species in response to ABA. Left, confocal images of whole-mount HCR showing the detection of OG03053 orthologs across species. Scale bars, 50 μm. Right, average signal intensity of OG03053 orthologs across defined cell type groups.



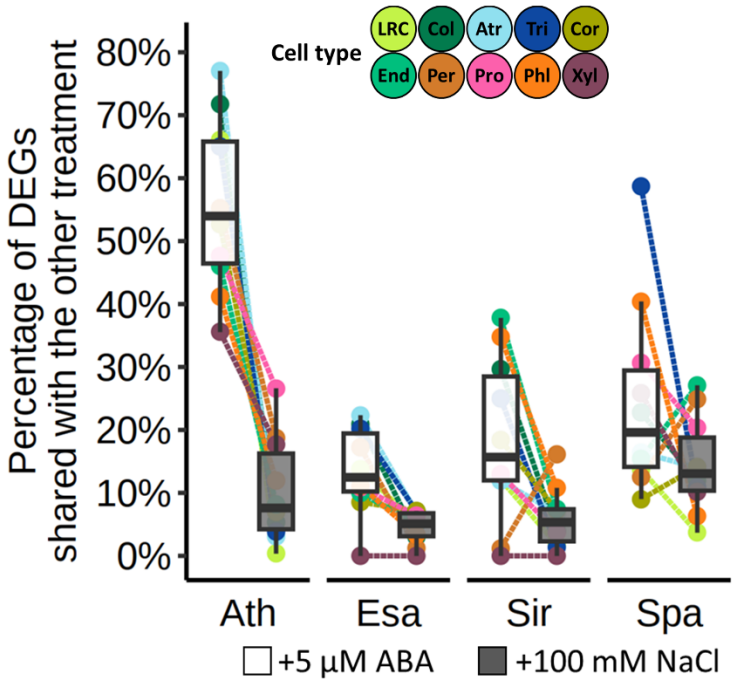
Supplementary Fig. 22. Expression changes of OG10873 orthologs (*AT5G47450*) across species in response to NaCl. a Expression of OG10873 orthologs across species from scRNA-seq data under different conditions. Left, UMAP representation of the expression of OG10873 orthologs. Right, pseudo-bulk cell type expression of OG10873 orthologs. b Whole-mount *in situ* HCR detection of expression changes of OG10873 orthologs across species in response to ABA. Left, confocal images of whole-mount HCR showing the detection of OG10873 orthologs across species. Scale bars, 50 μm. Right, average signal intensity of OG10873 orthologs across defined cell type groups.



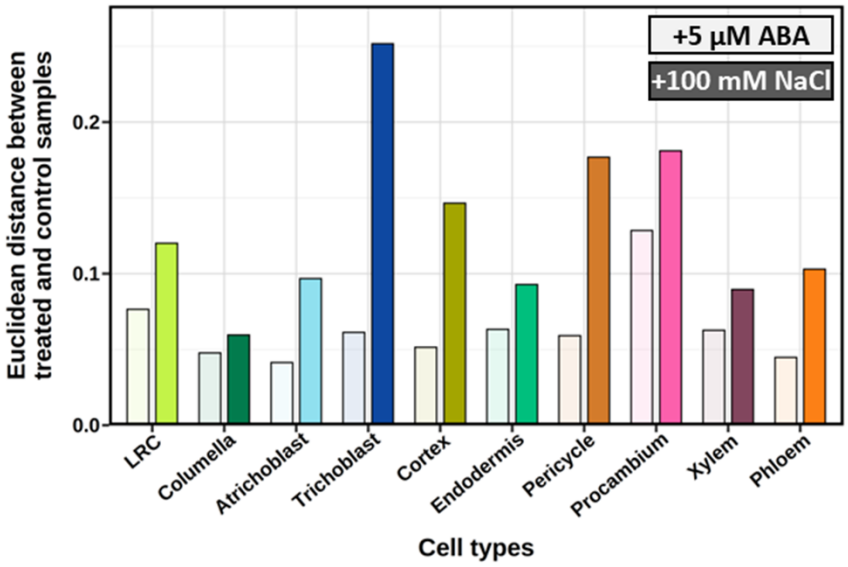
Supplementary Fig. 23. Number of stress responsive genes in each cell type from each species in response to 5 µM ABA and 100 mM NaCl. Total number of stress responsive genes (nr, non-redundant) from all cell types in each species was shown on the top left corner.



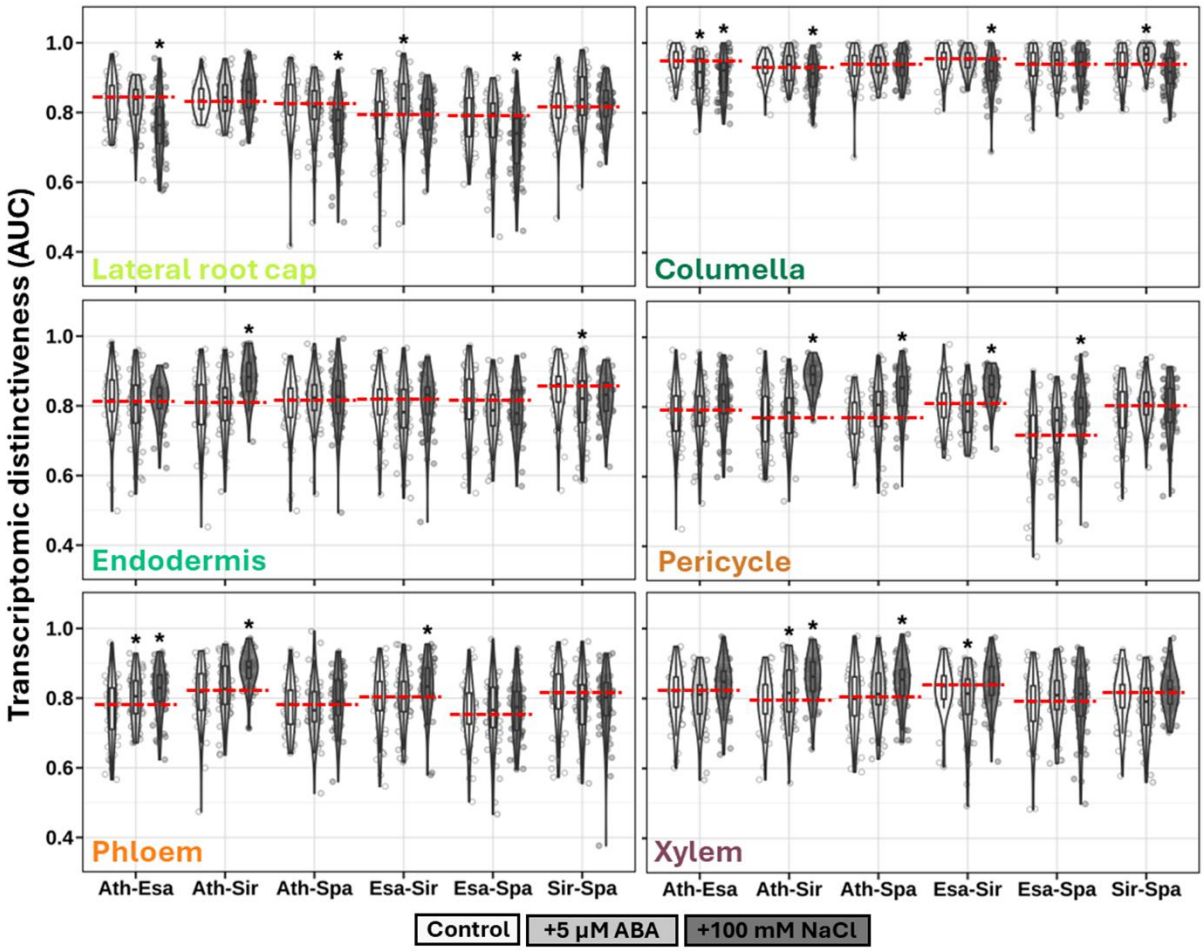
Supplementary Fig. 24. Overlap of stress responsive genes in each developmental zone from each species in response to 5 µM ABA (left) and 100 mM NaCl (right).



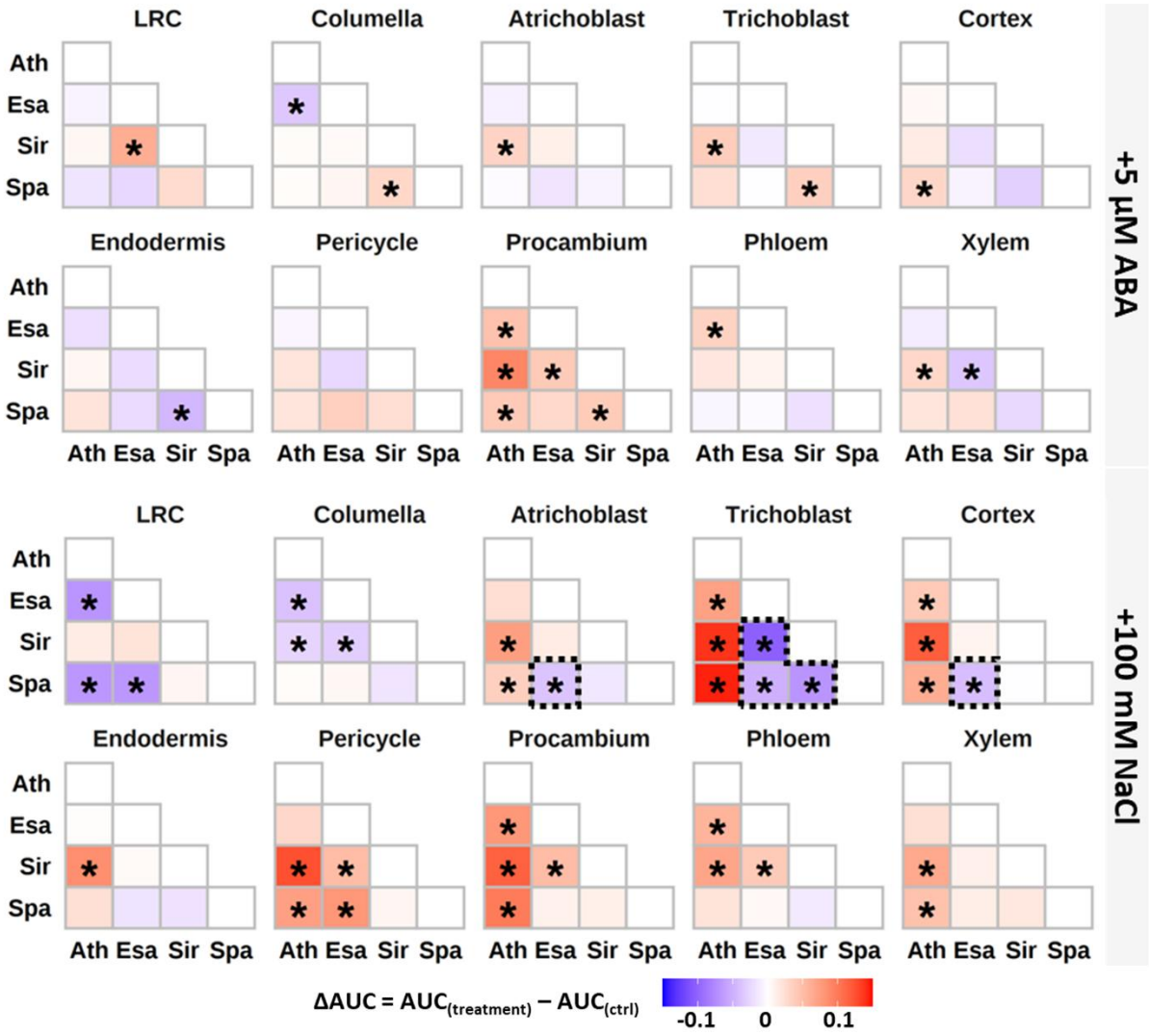
Supplementary Fig. 25. The percentage of DEGs shared between ABA and NaCl treatment. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Sir, *Sisymbrium irio*; Spa, *Schrenkiella parvula*.



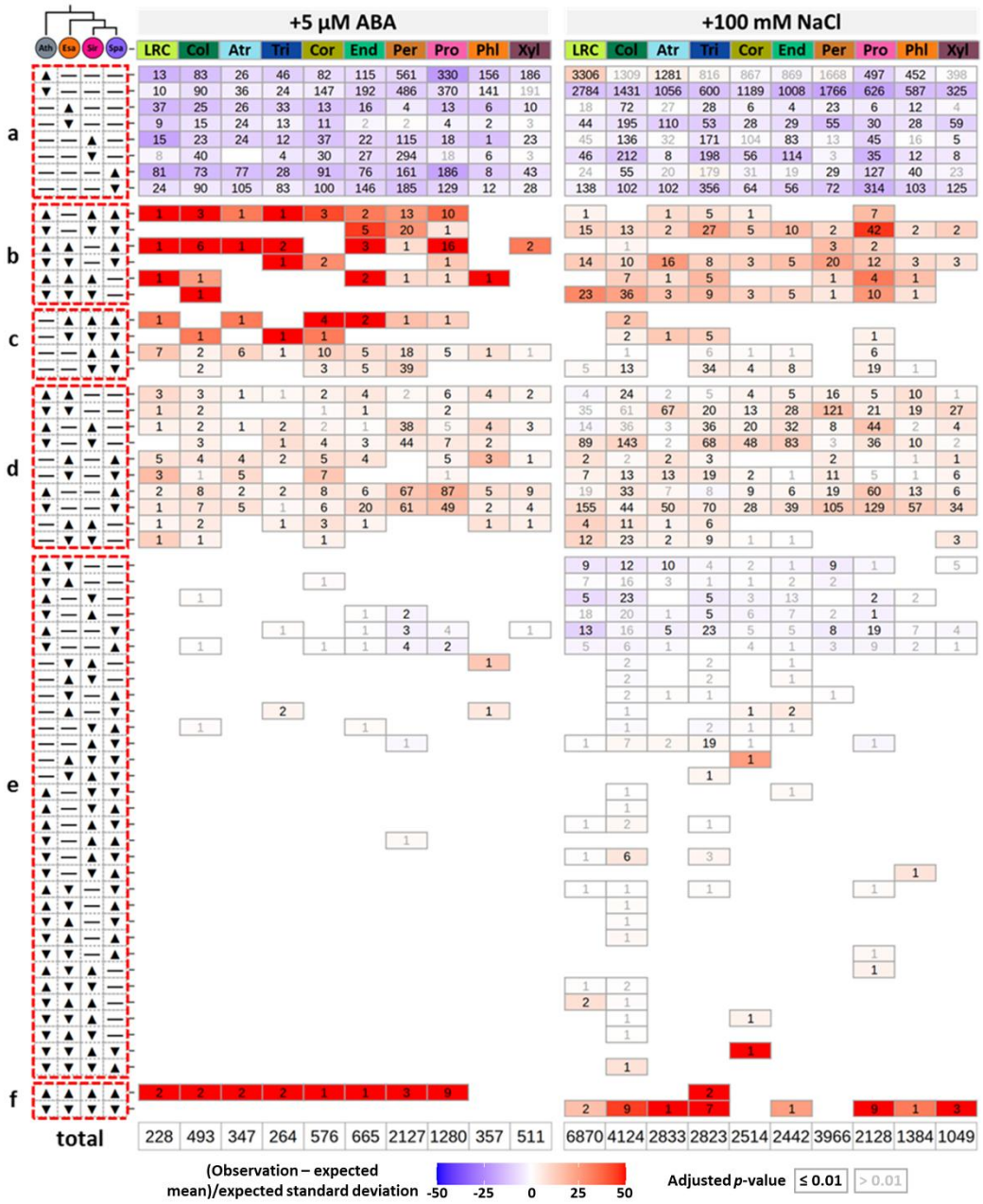
Supplementary Fig. 26. Euclidean distance between treated and control samples in Fig. 5e. The salt-treated samples were generally more distant (greater Euclidean distances) to the control samples than the ABA-treated samples.



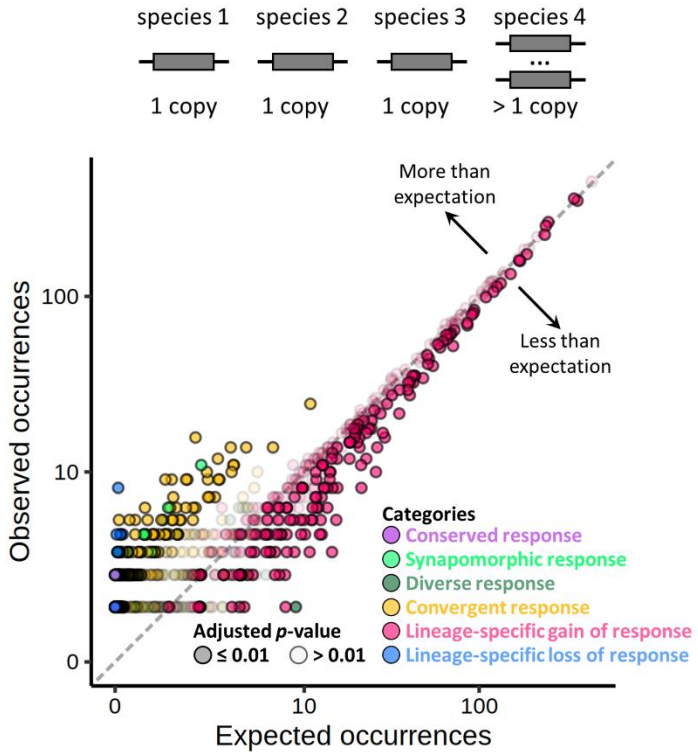
Supplementary Fig. 27. Transcriptomic distinctiveness between species pairs under different conditions. AUC here quantifies the separability between species for each cell type. AUC = 0.5 indicates random separability; AUC = 1.0 indicates perfect separability. Asterisks indicate significance after comparing to control (n = 50) with Student's t-test at *p*-value cutoff of 0.05. The red dash lines mark the medians the transcriptomic distinctiveness under control condition. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Sir, *Sisymbrium irio;* Spa, *Schrenkiella parvula*.



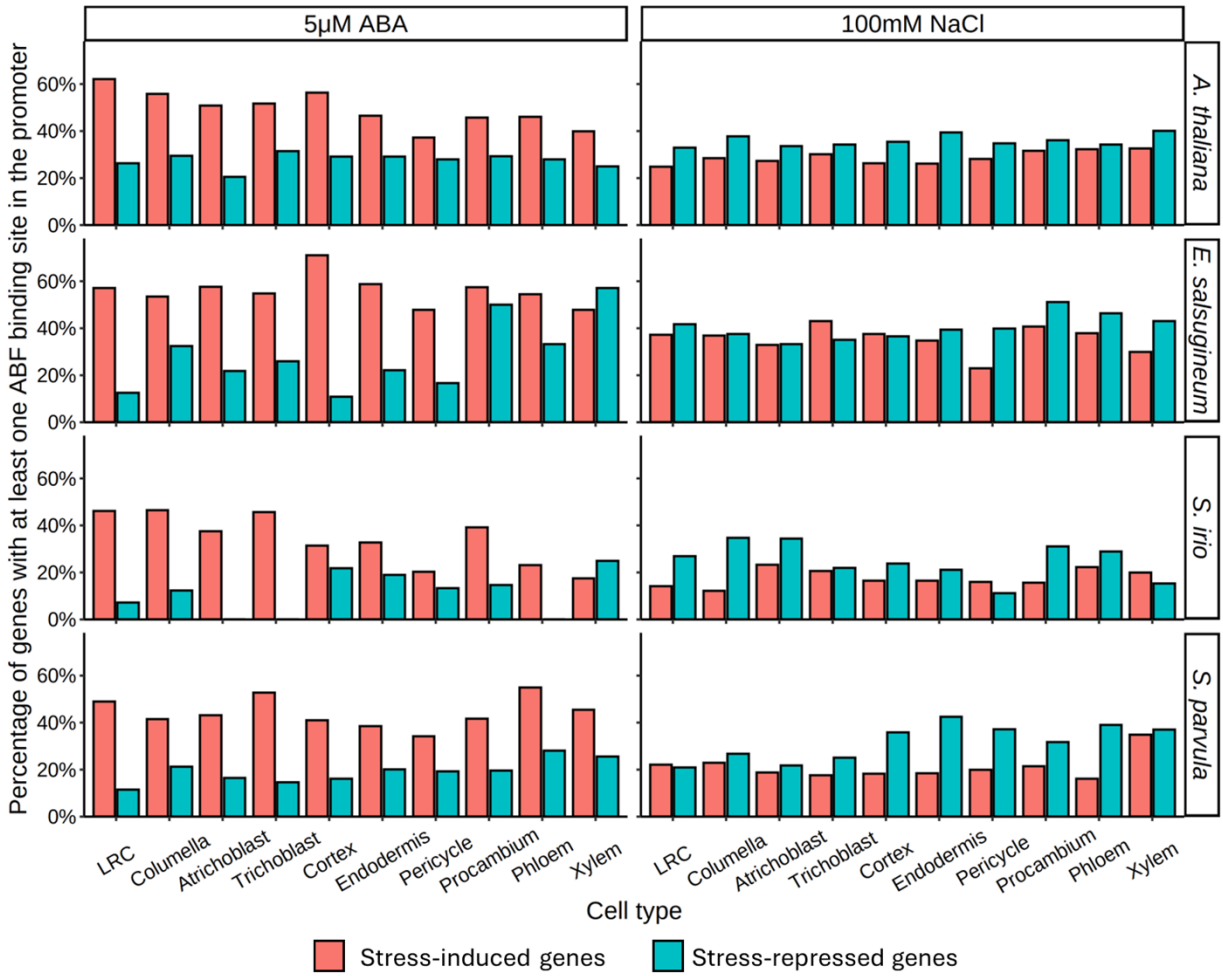
Supplementary Fig. 28. Changes in cell type transcriptome divergence between species in response to ABA (top) and NaCl (bottom). The darker colors indicate larger differences in transcriptome divergence induced by the treatment. Asterisks indicate significance after comparing to control (n = 50) with Student's t-test at p-value cutoff of 0.05. The black boxes highlight the comparisons of interest. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Sir, *Sisymbrium irio*; Spa, *Schrenkiella parvula*.



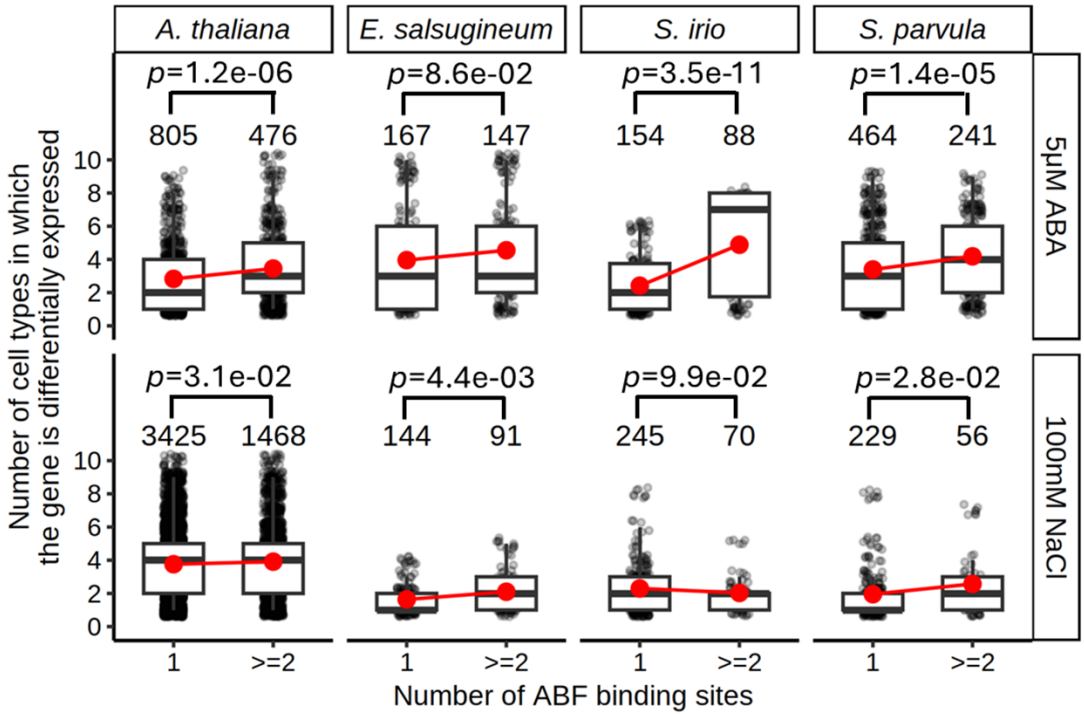
Supplementary Fig. 29. Differences between observed and bootstrapped (10,000 times) occurrences of different ortholog response patterns. The patterns were grouped according to orthologs showing a, lineage-specific gain of response; b, lineage-specific loss of response; c. synapomorphic response; d. convergent response; e, diverse response; f, conserved response. Number of orthologs for each pattern in each cell type is indicated in the corresponding tile. The differences were measured as the number of standard deviations between the overserved occurrences and the corresponding bootstrapped means and indicated with the shade of the color with darker colors corresponding to larger differences. Black and grey text represent significant and insignificant deviation of the observed occurrences from the expected distribution, respectively, at Benjamini-Hochberg-adjusted p-value ≤ 0.01.



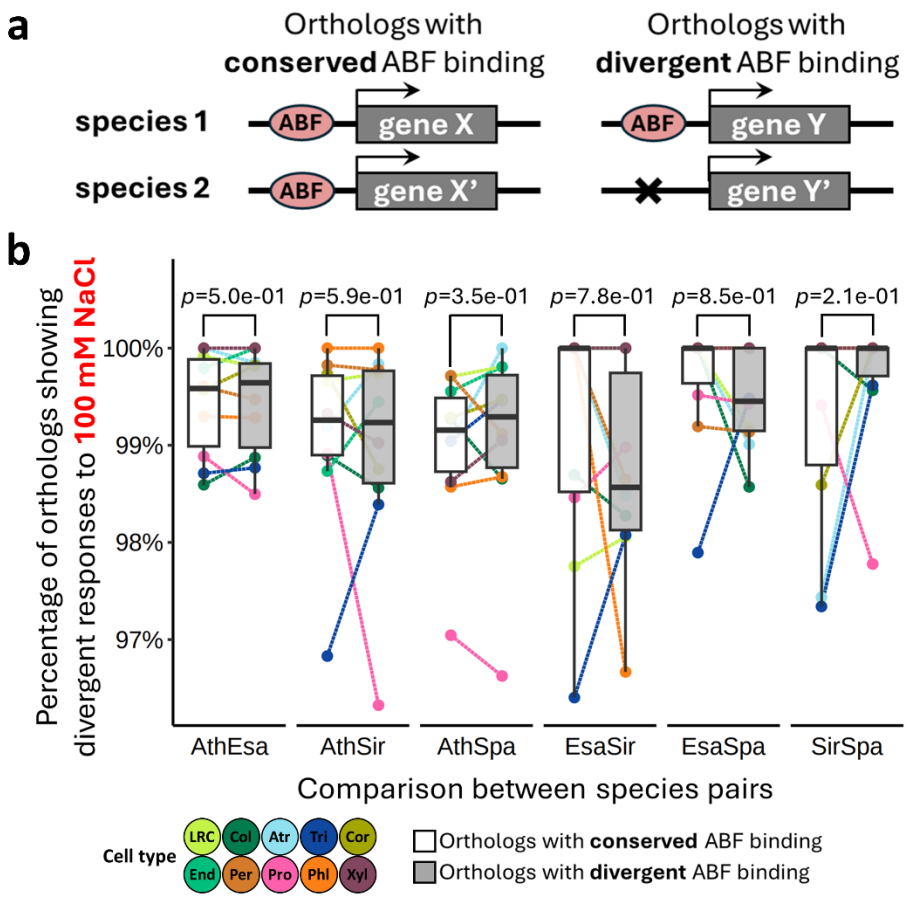
Supplementary Fig. 30. Differences between observed and bootstrapped (10,000 times) occurrences of response patterns among all ortholog groups that have been duplicated only in one species. The darker and lighter color intensities represent significant and insignificant deviation of the observed occurrences from the expected distribution, respectively, at Benjamini-Hochberg-adjusted p-value ≤ 0.01.



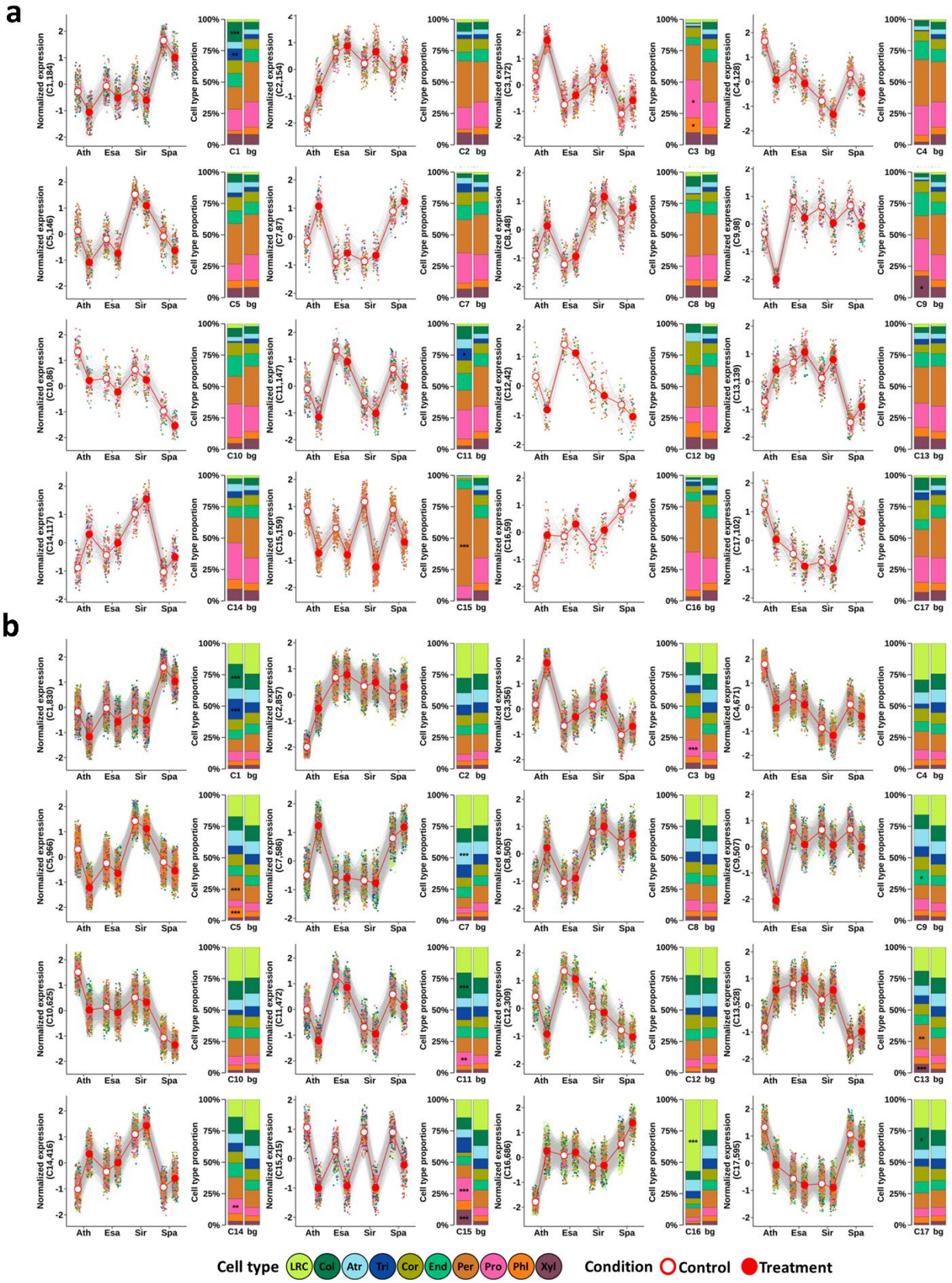
Supplementary Fig. 31. Percentage of genes with ABF binding among genes induced by 5μM ABA (left) or 100mM NaCl (right). ABF binding was defined as the presence of ABF binding events in the 2-kilobase upstream of the transcription start site using the DAP-seq data from Sun *et al.*, 2022[9].



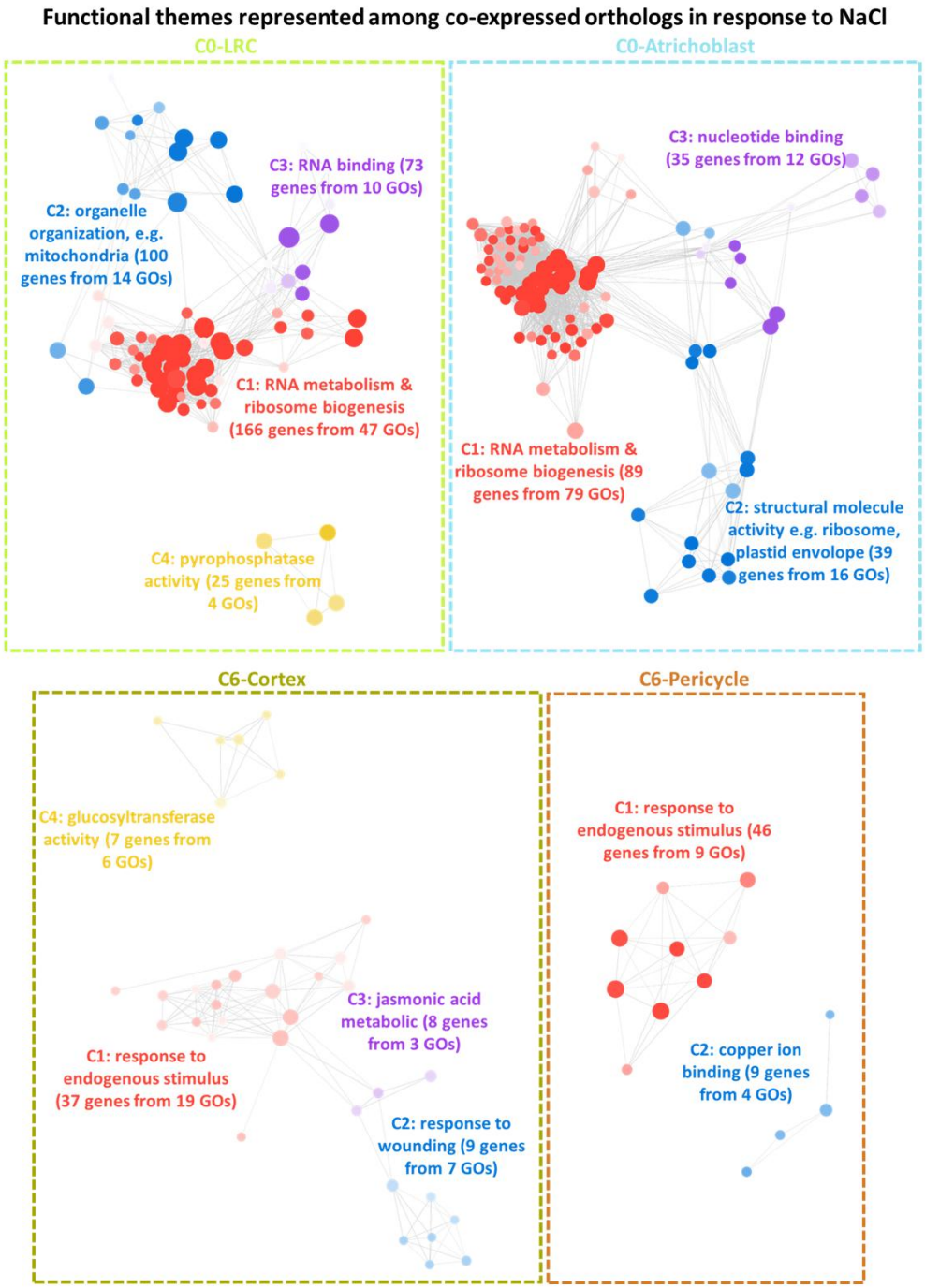
Supplementary Fig. 32. Number of responsive cell types for genes with different numbers of ABF binding sites. ABF binding was defined as the presence of ABF binding events in the 2-kilobase upstream of the transcription start site using the DAP-seq data from Sun *et al.*, 2022[9]. Significance was represented by *p*-values determined by Student's t-test and shown on the top of each comparison. Numbers above each box indicate the number of genes included in each group.



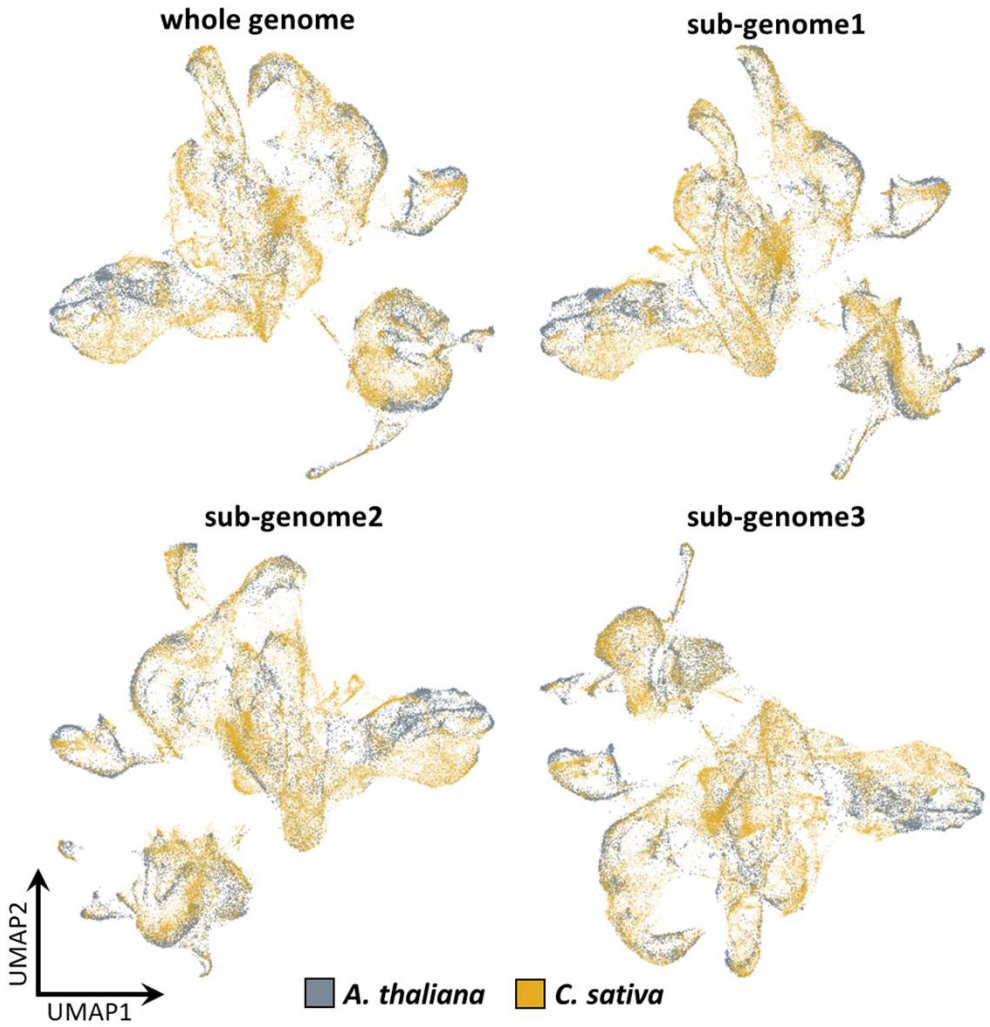
Supplementary Fig. 33. The relationship between ABF binding and stress responses of orthologs between species. a Categorization of orthologs with different ABF binding events. ABF binding was defined as the presence of ABF binding events in the 2-kilobase upstream of the transcription start site using the DAP-seq data from Sun *et al.*, 2022[9]. b Percentage of divergently regulated orthologs (in response to 100 mM NaCl) with conserved or divergent ABF binding. Significance was represented by *p*-values determined by Wilcoxon test and shown on the top of each comparison.


a
b
Normalized expression (C1,184)
Normalized expression (C2,154)
Normalized expression (C3,172)
Normalized expression (C4,128)
Normalized expression (C5,146)
Normalized expression (C7,87)
Normalized expression (C8,148)
Normalized expression (C9,98)
Normalized expression (C10,86)
Normalized expression (C11,147)
Normalized expression (C12,42)
Normalized expression (C13,139)
Normalized expression (C14,117)
Normalized expression (C15,159)
Normalized expression (C16,59)
Normalized expression (C17,102)
Normalized expression (C1,830)
Normalized expression (C2,857)
Normalized expression (C3,356)
Normalized expression (C4,671)
Normalized expression (C5,966)
Normalized expression (C7,586)
Normalized expression (C8,505)
Normalized expression (C9,507)
Normalized expression (C10,625)
Normalized expression (C11,472)
Normalized expression (C12,309)
Normalized expression (C13,528)
Normalized expression (C14,416)
Normalized expression (C15,215)
Normalized expression (C16,686)
Normalized expression (C17,595)
Cell type proportion
Ath Esa Sir Spa
C1 bg C2 bg C3 bg C4 bg
C5 bg C7 bg C8 bg C9 bg
C10 bg C11 bg C12 bg C13 bg
C14 bg C15 bg C16 bg C17 bg
Cell type LRC Col Atr Tri Cor End Per Pro Phl Xyl
Condition Control Treatment

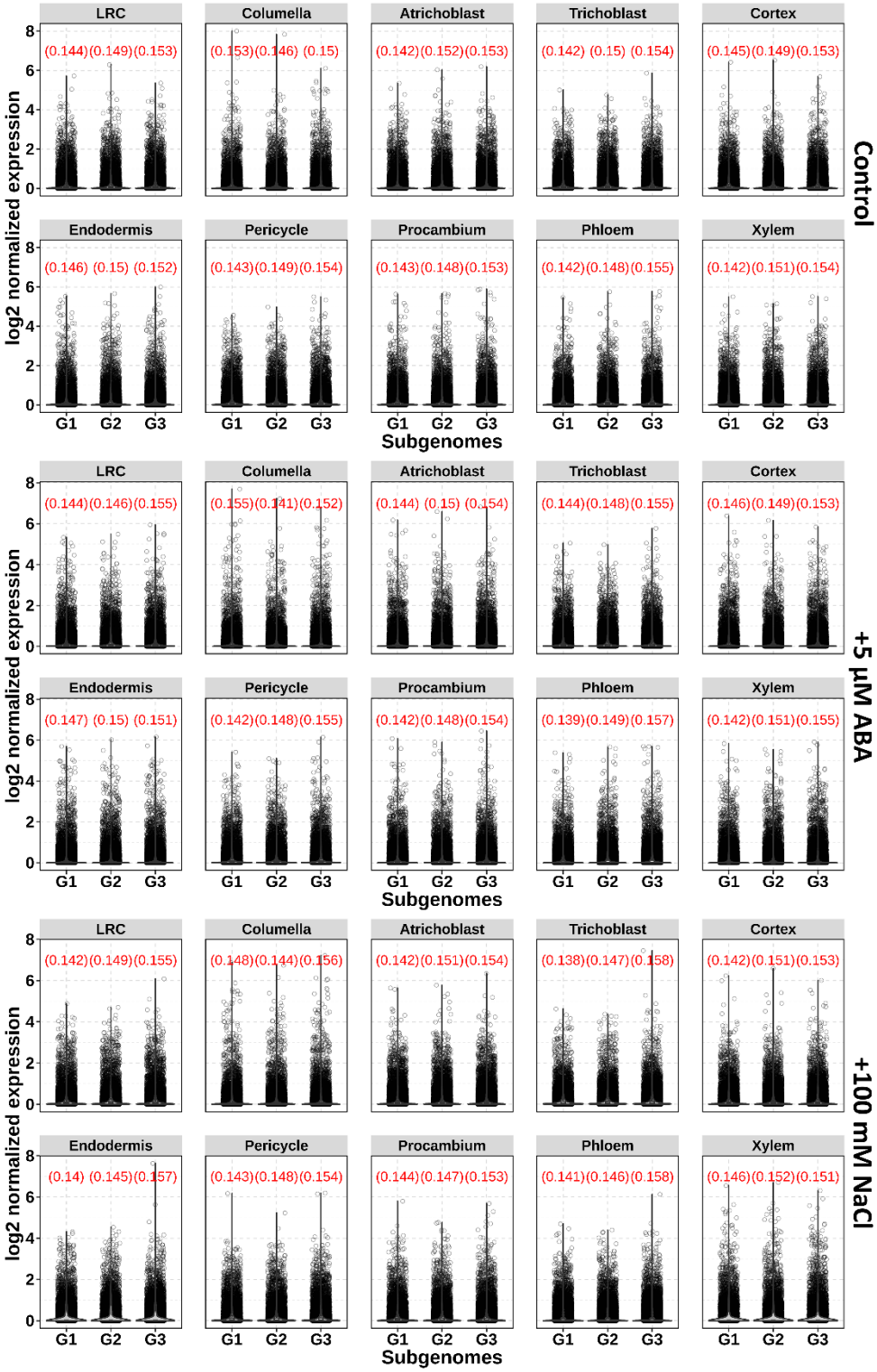
Supplementary Fig. 34. Contributions of each cell type to co-expression networks identified from stress-responsive orthologs under 5 µM ABA (a) and 100 mM NaCl (b). Cluster ID and the number of orthologs in the cluster are shown in parentheses in the Y-axis labels. Cell type proportions of the cluster were compared to background (bg) where orthologs from all clusters were averaged (see Methods for details). Asterisks represent significance determined by hypergeometric test, * p-value < 0.01, ** p-value < 0.001, *** p-value < 0.0001. LRC, Lateral Root Cap; Col, Columella; Atr, Atrichoblast; Tri, Trichoblast; Cor, Cortex; End, Endodermis; Per, Pericycle; Pro, Procambium; Phl, Phloem; Xyl, Xylem. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Sir, *Sisymbrium irio;* Spa, *Schrenkiella parvula*.



Supplementary Fig. 35. Functional representations of co-expressed orthologs in response to NaCl. GO clusters are differently colored and labelled with the representative functional term. Each node represents a gene ontology (GO) term; node size represents genes in the test set assigned to that functional term; GO terms sharing > 50% of genes are connected with edges; and the shade of each node represents the *p*-value assigned by the enrichment test at adjusted *p* values < 0.05 with darker shades indicating smaller *p*-values.

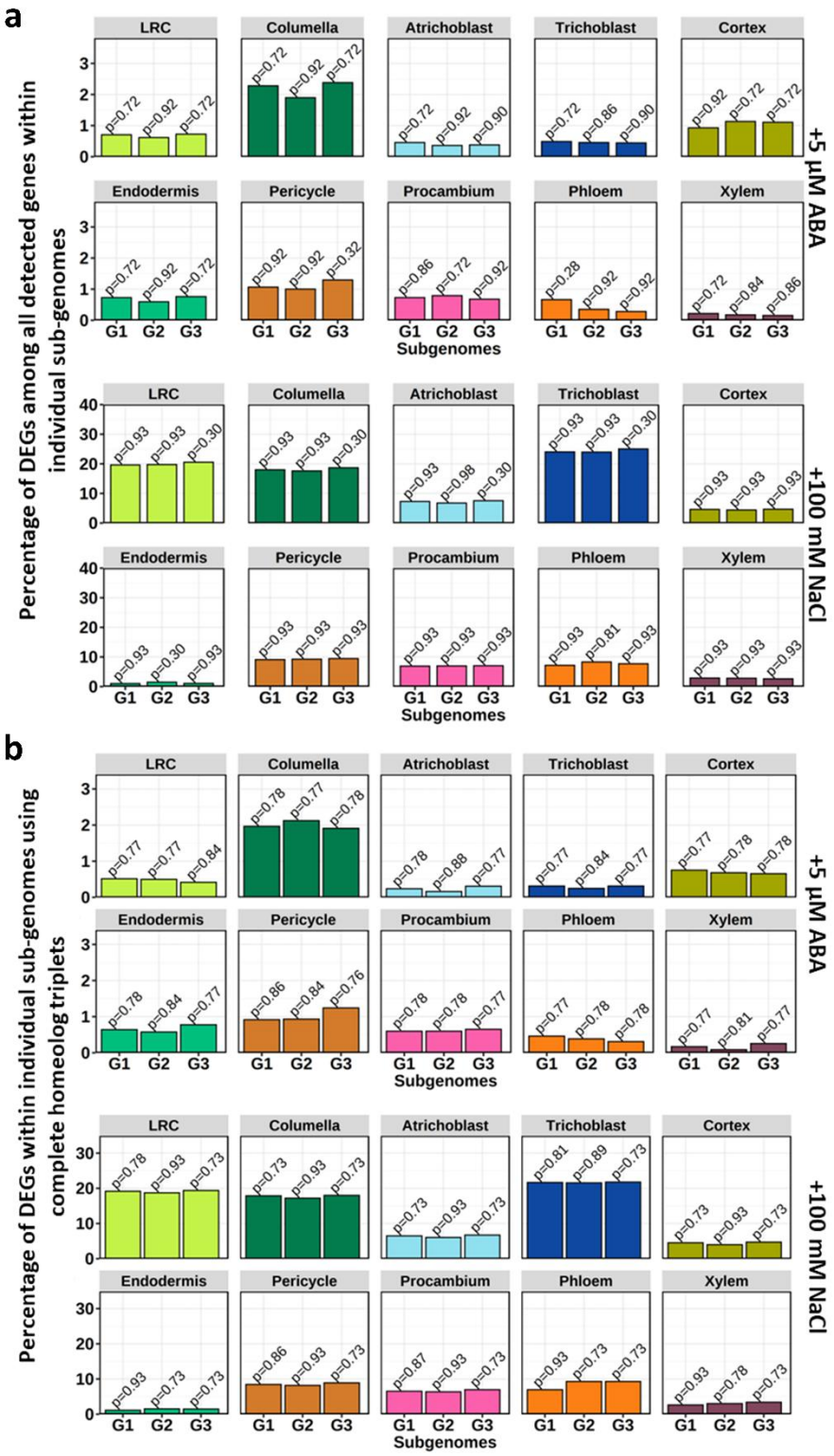


Supplementary Fig. 36. Integration between Arabidopsis and *C. sativa* using different sub-genomes in *C. sativa*.



Supplementary Fig. 37. Overall transcriptome expression levels separated by the three sub-genomes for each cell type under different conditions. The number of genes considered for sub-genome G1, G2, and G3 is 21,848, 21,682, and 23,066, respectively. Numbers in red indicate the averages of transcript levels.

Supplementary Fig. 38. Comparison of GO functional enrichment among DEGs in response to ABA (top) and NaCl (bottom) for each cell type across species. The enrichment in *C. sativa* was performed on DEGs combined from all sub-genomes. Endodermis, phloem and xylem cells in *C. sativa* were removed from the comparison for NaCl treatment due to insufficient number of cells (< 50 cells). The detailed list of GO terms is present in Supplementary Data 6. The shade represents the *p*-value assigned by the enrichment test at adjusted *p*-value < 0.05 with darker shades indicating smaller *p*-values.



Supplementary Fig. 39. Percentage of DEGs in each sub-genome of *C. sativa*. The percentage for each genome was calculated considering all detected genes within individual sub-genomes (a) and only the complete homeolog triplets (b). Numbers on top of the bars represent adjusted *p*-values from hypergeometric test.