Diversification of gene expression across extremophytes and stress-sensitive species in the Brassicaceae

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

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

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

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

[4]DOE Joint Genome Institute, Berkeley, California 94720, USA

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

[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

*These authors contributed equally to this work.

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

SUMMARY

Stress-sensitive and stress-adapted plants respond differently to environmental stresses. To explore the cellular-level stress adaptations, we built root single-cell transcriptome atlases for diverse Brassicaceae species: stress-sensitive plants (*Arabidopsis thaliana* and *Sisymbrium irio*), extremophytes (*Eutrema salsugineum* and *Schrenkiella parvula*) and a polyploid crop (*Camelina sativa*), under control, NaCl, and abscisic acid treatments. Approximately half of Arabidopsis cell-type markers lacked expression conservation across species. We identified new conserved cell-type markers, along with orthologs showing divergent expressions. We experimentally mapped distinct cortex sub-populations to different cortex layers across species. We found distinct cell-type-specific transcriptomic responses between species and treatments. Lineage-specific losses of stress responses were less prevalent but evolutionarily more favored than gains. In *C. sativa*, sub-genomes contributed equally to stress responses and homeologs with divergent stress responses typically did not exhibit high coding sequence or expression divergence. Our study provides a foundational root atlas and an analytical framework for multi-species single-cell transcriptomics.

KEYWORDS

extremophytes, oil crop, scRNA-seq, comparative transcriptomics, ABA, salt-stress adaptation, ortholog expression, root development, homeolog evolution in polyploids

INTRODUCTION

Environmental stress responses are complex traits that require spatiotemporal coordination of gene networks acting in diverse assemblage of cell types. Plant roots play a pivotal role in responding to edaphic stresses, having evolved as the primary interface with soil. Roots are a mixture of functionally and transcriptionally diverse cell types, each perceiving and responding to environmental stresses in a development-dependent way (Dinneny et al., 2008; Iyer-Pascuzzi et al., 2011; Kawa & Brady, 2022). Hence, single-cell transcriptomics provides a means to dissect the developmentally coordinated stress responses at a cellular level that synergistically define the response at the whole plant level.

The developmental and stress responsive gene networks identified in the model plant *Arabidopsis thaliana* (Waadt et al., 2022; Zhang et al., 2020) provide only partial insights into the diverse evolutionary trajectories employed by plants to cope with environmental stresses. This limitation becomes more evident when considering the rewired and novel stress tolerance mechanisms employed by naturally stress-adapted species such as extremophytes that exhibit a remarkable ability to optimize growth and survival under challenging environmental conditions (Xu et al. 2020). Therefore, investigation of stress-responsive gene networks in extremophytes provides opportunities to uncover evolutionary innovations that confer stress resilience. Establishment of a comparative approach that takes advantage of our understanding in model systems and the diversity present in the plant kingdom thus enhances our understanding of the diversity of stress adaptations in plants, but also strengthens our capacity to source the most effective genetic strategies for crop improvement amidst a climate crisis where the intensity and frequency of environmental stresses are becoming increasingly unpredictable.

In this study, we have selected the extremophyte models, *Eutrema salsugineum* and *Schrenkiella parvula*, the more stress-sensitive models, *Arabidopsis thaliana* (hereinafter referred to as Arabidopsis) and *Sisymbrium irio*, and the polyploid oil crop, *Camelina sativa* in the Brassicaceae family. These species exhibit divergent physiological responses not only to high salinity, but also to the stress hormone, abscisic acid (ABA) (Kazachkova et al. 2018; Sun et al. 2022). When exposed to salt stress, both extremophytes exhibit greater salinity tolerance than the non-extremophytes, yet their transcriptional responses differ significantly from one another (Tran et al., 2022). Their responses to external ABA treatments are similarly diverse with *S. parvula* being drastically distinct from the other species (Sun et al., 2022). The underlying molecular mechanisms that enable the observed divergences in stress responses among these

species have so far been generalized to whole root and shoot tissues. However, how different cell populations collectively give rise to the organ-level response remains largely unknown.

We generated primary root transcriptome atlases at single-cell resolution for five Brassicaceae species under stress-neutral as well as ABA- and salt-treated conditions. We developed an approach and methodology for cross-species cell-type annotation and comparative single-cell transcriptomics that can be readily transferred to other multi-species comparisons. Deployment of these tools allowed us to identify evolutionary innovations in gene expression patterns; confirm the existence of developmentally coordinated, species-specific cell states; determine the uneven contribution of cell types to overall stress responses among species; describe the capacity of sub-genomes to rapidly gain diverse expression profiles in a polyploid crop; and reveal gene networks with modified functions associated with stress responses. Together, this work establishes a foundational root cell atlas resource to explore cell-specific, species-specific, or stress-dependent mechanisms across ecologically diverse Brassicaceae model plants.

RESULTS

The expression of nearly half of the detected Arabidopsis cell type markers was not conserved across species

We generated 36,000 to 50,000 root cell transcriptomes for each of the target diploid species (Arabidopsis, *E. salsugineum*, *S. irio*, and *S. parvula*) under control (stress-neutral), 5 µM ABA, and 100 mM NaCl treatments using the 10X Genomics scRNA-seq platform to provide over 174,000 root cell transcriptomes in total (Figure 1 and S1). We used ABA as a more generic abiotic stress stimulus and NaCl as a more specific stimulus, and the concentrations used were informed by our previous studies to ensure significant responses but not irreversible tissue damage (Oh et al., 2014; Sun et al., 2022; Tran et al., 2022). To facilitate the utility of this multi-species dataset as a community resource, we generated an accompanying interactive web browser to enable exploration of this atlas at https://plantbiology.shinyapps.io/brassicaceaeatlas/. Preprocessed data matrices partitioned by cell types are provided in the searchable browser for convenient reanalysis as well.

Root cell-type marker genes have been established for Arabidopsis (Coate et al., 2020; Denyer et al., 2019; Jean-Baptiste et al., 2019; Liu et al., 2021; Ryu et al., 2019; Shahan et al., 2022; Wendrich et al., 2020), though their applicability as markers for root cell types in related species has been untested. We developed a two-step strategy to annotate cell types for Arabidopsis and non-Arabidopsis cells in our dataset (Figure S2; Methods). First, we built a reference atlas using Arabidopsis cells from the control condition, and annotated individual cells based on the collective outcome of three independent methods: correlation-based annotation (where individual cells were annotated based on their transcriptomic similarity to reference cell type transcriptomes), marker-based annotation (where individual cells were annotated based on the preferential expression of known cell type markers), and integration-based annotation (where

individual cells were annotated based on their mapping to reference single cell datasets). Second, we employed three prediction algorithms (TransferData in Seurat(Stuart et al., 2019), Symphony (Kang et al., 2021), and scArches (Lotfollahi et al., 2022)) to transfer Arabidopsis annotations to individual cells in non-Arabidopsis species based on the 15,198 unambiguous 1-to-1 orthologous genes identified (details in Methods). This resulted in the identification of ten major root cell types for all four species in our dataset (visualized on the integrated UMAP space, Figure 1A-C). We did not detect a recognizable cell population for quiescent center (QC) cells in our dataset (Figure S3), likely due to under-sampling of this rare cell type. Our overall cell-type identification indicates that these diverse Brassicaceae species do not differ in the presence of the major root cell types.

The expression fidelity of known Arabidopsis cell-type marker genes was examined in the other target species (Figure 1D and S4). Out of 461 previously established Arabidopsis cell-type markers, we found 1-to-1 orthologs for 254 of them, 241 of which were representative marker genes for the ten major root cell types described above (Table S1). Many of these marker orthologs showed enriched expression in the expected cell types across species (Figure 1D and S4). However, only 144 of these Arabidopsis markers had expression levels above the threshold of detection in our Arabidopsis data (Figure 1E). Among these remaining Arabidopsis markers, 48% (69/144) orthologs lost expression specificity in one or more species (Figure 1D and E). The lowest rate of conservation for these Arabidopsis cell-type markers was observed in *E. salsugineum* (Figure 1E and S4). The cell-type markers for vascular tissues are generally highly conserved across species, while markers from outer cell layers of the root, especially the root cap and epidermal cells, were least conserved (Figure S4).

To obtain an improved set of Brassicaceae cell type markers, we utilized our multi-species scRNA data to identify orthologous genes exhibiting highly conserved preferential expression for each of the ten cell types across these species (Figure S5; Table S2). To test the accuracy of these new cell-type markers, we optimized a whole-mount hybridization chain reaction (HCR)-based *in situ* hybridization method (Choi et al., 2018) for *in vivo* mapping of the selected markers across species. *In situ* expression patterns of the selected markers matched the predicted cell-type specificity based on the scRNA-seq data (Figure S6). This provided independent support for the cell-type assignments in the target species and shows that these orthologous genes represent broadly conserved novel 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 (Figure 2A and S7). The overall correlations of cell-type expression among orthologs were modest even between each pair of species (Figure 2A). To determine the patterns of divergence in cell-type expression between species under stress-neutral conditions, we compared all 1-to-1 ortholog expression across the four species in a phylogenetic framework, and detected 350 orthologs showing diverged expression profiles in one phylogenetic branch compared to others or between extremophytes

and non-extremophytes (Figure 2B; Table S3). For example, there were 78 orthologs whose expression across cell types in Arabidopsis (lineage I) differed from the cell-type expression of the orthologs in the other three (lineage-II) species. We also found 43 orthologs whose cell-type expression differentiate extremophyte and non-extremophyte species (Figure 2B), independent of their phylogenetic relationships.

We examined two of these cell-type expression profile changes using HCR. Both scRNA-seq data and HCR results showed that the gene *NIP5;1*, which encodes a plasma membrane-localized aquaporin, is predominantly expressed in the outer cell layers, from epidermal to endodermal cells, in Arabidopsis, *S. irio*, and *E. salsugineum*, while its expression is restricted to endodermal cells in *S. parvula* (Figure 2C). Arabidopsis AtNIP5;1 serves as a channel responsible for transporting boron into roots (Takano et al., 2006). *S. parvula* is native to boron-rich soils in Anatolia, Turkey and can tolerate high boron levels (Oh et al., 2014; G. Wang et al., 2021) and the restricted expression of *NIP5;1* in *S. parvula* may reduce boron uptake into the vascular stream, thus contributing to *S. parvula*'s unique tolerance to boron toxicity.

Our scRNA-seq data indicate that *GSTU25*, encoding a plant-specific tau class glutathione S-transferase, is preferentially expressed in the root cap of non-extremophytes. In the extremophytes, its expression becomes enriched in the endodermal cells (Figure 2D). HCR experiments confirmed the absence of endodermal expression of *GSTU25* in the non-extremophytes (Figure 2D). However, our scRNA-seq data did not fully agree with the HCR observations, particularly for *E. salsugineum* where HCR, but not the scRNA-seq data, detected the presence of *GSTU25* transcripts in epidermal and cortex cells (Figure 2D). This discrepancy may be due to the technical differences between the two complementary RNA detection methods. Glutathione S-transferases protect cells from oxidative damage by utilizing glutathione to scavenge reactive oxygen species (ROS) (S. Kumar & Trivedi, 2018). *GSTU25* has been reported to be induced under oxidative, salt, and heat stresses (Sappl et al., 2009; Terrón-Camero et al., 2022). The convergent endodermal expression of *GSTU25* in the two extremophytes might suggest a role of *GSTU25* in ROS scavenging in endodermal cells, which may contribute to the salt tolerance in the two extremophytes.

Distinct cortex cell subgroups support lineage-specific root anatomy

Anatomical analysis of *S. irio* and *S. parvula* indicates that they possess two cortical cell layers, whereas Arabidopsis and *E. salsugineum* possess a single cortical cell layer at the early developmental stages of the seedling (Figure S8A). Agreeing with this observation, UMAP-based analysis of cortex cells identified two major cortex subpopulations from Arabidopsis and *E. salsugineum*, while three cortex subpopulations are found in *S. parvula* and *S. irio* (Figures S8B). We hypothesized that the extra cortex cell population identified in *S. irio* and *S, parvula* is derived from the additional cortex layer. Under this hypothesis, we would expect that the extra cortex cell population should be underrepresented in Arabidopsis and *E. salsugineum* but not in *S. parvula* and *S. irio*. To test this, we selected the cortex subpopulation structure from *S. irio* which showed the most separated clusters as the reference and categorized the three subpopulations from this species into left, middle, and right subgroups (Figure 3A). Cortex cells

from the other three species were mapped to the *S. irio* cortex subpopulations based on 1-to-1 orthologs (Figure 3B). While some representation of the three cortex subpopulations was detected in all species (Figures 3B and C), the middle population was highly depleted in Arabidopsis and *E. salsugineum*, but not in *S. parvula* (Figures 3C). To confirm that the middle cortex subpopulation represented the extra cortex layer found in *S. irio* and *S, parvula*, we identified marker genes that were preferentially expressed in the middle population (Figure S9A Figure 3D; Table S4) and inspected their spatial expression in roots. The *in situ* hybridization results showed that *Si-s1384-01650*, which was specifically expressed in the middle population (Figure 3D, upper panel), distinctly marked the inner cortex in *S. irio* (Figure 3D, lower panel), in support of our hypothesis. The orthologs of *Si-s1384-01650* in the other three species also showed consistent results with their expression pattern predicted by the scRNA-seq data (Figure 3E).

To further investigate the transcriptomic differences between the inner and outer cortex layers in *S.irio*, we compared the transcriptome of the inferred inner and outer cortex layers (left and right subpopulations, Figure S9-S11). This identified several cellular processes that were differently partitioned between the two layers (Figure S9), including flavonoid biosynthesis which has been previously shown to be preferentially localized to cortex cells (Daryanavard et al., 2023). We subsequently detected a strong expression bias towards the inner layer in *S. irio* for the majority of the genes involved in flavonoid biosynthesis, including CHALCONE SYNTHASE (CHS) (Figure 3F, left panel), which catalyzes the first committed step in this biosynthetic pathway. Intriguingly, this inner cortex layer-bias is also conserved in *S. parvula* cortex cells (Figure 3F, right panel). Together these data demonstrate transcriptional divergence in the cortex cell populations associated with the increase in anatomical complexity of *S. irio* and *S. parvula* roots and suggest that this may have led to the specialization of the inner layer for flavonoid metabolism.

ABA and NaCl induce distinct cell-type-specific responses

To obtain an overview of cell-type transcriptional diversity across species and conditions, we performed a principal component analysis (PCA) based on pseudo-bulk cell-type transcriptomes (Figure 4A). 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 on the plane of the largest two PCs. Given that *S. parvula* and *S. irio* share a closer phylogenetic relationship compared to the two extremophytes, this suggests that the extremophytes share a portion of similarities in their cell-type specific transcriptomes beyond the phylogenetic relationships. Within each species, outer root cell layers (root cap, epidermal, cortex, and endodermal cells) displayed distinct transcriptional signatures in contrast to inner cell layers (pericycle, procambium, phloem, and xylem cells), regardless of treatments. This divergence in transcriptomes between the inner and outer cell layers in the root suggests the presence of distinct signaling environments within these two domains of the root.

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; Table S5). Consistent with prior findings (Sun et al. 2022; Tran et al. 2022), Arabidopsis had the most pronounced responses to the ABA and NaCl treatments among the four species. This was evident in the higher number of differentially expressed genes (DEGs), ranging from 2 to 10 times more than the other species (Figure S12). However, within each species, the degree of response varied significantly across cell types and treatments. To further compare cell type responsiveness to treatments, we implemented Augur (Skinnider et al., 2021), a method to rank cell types based on their magnitude of response to biological perturbations in single-cell data. Procambium was the cell type most responsive to ABA in Arabidopsis and *E. salsugineum* (Figure 4B). In contrast, columella cells were most responsive to ABA in *S. irio* and *S. parvula* (Figure 4B). The cell-type stress responses further diverged into species specific responses when plants were treated with 100 mM NaCl. While the most responsive cell types largely came from the inner cell populations in Arabidopsis and *E. salsugineum*, the most dramatic changes were observed in root cap cells in *S. irio* and *S. parvula* (Figure 4B). While the two extremophytes' transcriptomes cluster closer to each other (Figure 4A), the transcriptional response to stress in each species depicts that *S. parvula* and *S. irio* are more similar, aligning with their closer phylogenetic relationship (Figure 4C).

To identify the most predominant cellular processes associated with stress response, we examined the functional processes enriched in DEGs (Figure 4D; Table S6). Response to stress was among the most commonly shared processes enriched in DEGs induced by ABA and DEGs repressed by NaCl treatments in the majority of cell types in all four species. RNA metabolism and ribosome biogenesis were found exclusively enriched among repressed genes in response to ABA. Additionally, they exhibited the most significant variation among cell-type specific responses across species. These processes were primarily enriched in inner cell types, like endodermal and stele cells, in Arabidopsis and *S. irio*, but in atrichoblast cells in *S. parvula* (Figure 4D). On the contrary, the same processes were enriched only among induced DEGs in response to the NaCl treatment across species. Moreover, these over-represented functions were primarily found in outer cell types up to pericycle cells in Arabidopsis, but restricted to only pericycle cells in *E. salsugineum* and epidermal cells in *S. parvula* (Figure 4D). These data suggest that similar biological functions are regulated by stress signaling in each species, but that their tissue specificity is what primarily diverges between species.

The observed divergence in stress responses among species seemed to primarily occur between cell types. Therefore, we computed alterations in transcriptomic distinctiveness between pairs of species for each cell type in response to ABA and NaCl treatments. This allowed us to pinpoint the cell types that exhibited the most pronounced response divergence. Under 5 µM ABA treatment, while most of the cell types remained largely transcriptomically similar across species, procambium cells showed increased transcriptomic divergence between species (Figure 4E). This suggests that distinct species-specific regulatory processes associated with ABA are predominantly active in procambium cells. Contrastingly, in response to 100 mM NaCl, we

observed an overall increase in response divergence among species from the root cap to inner cell layers (Figure 4E), indicating a collective contribution from different cell types to the diverging responses between species. Among them, trichoblast cells exhibited a strong phylogenetically aligned response where lineage I and II species responded distinctly to NaCl while atrichoblast and cortex cells presented increased similarity in transcriptomes after stress only between the two extremophytes (Figure 4E).

Distinct cell types contribute predominantly to the stress-prepared transcriptomes in the extremophytes

We aimed to investigate the occurrences of ortholog expression switching between stress-induced differential expression (DEGs) and stress-insensitive expression (non-DEGs), examining whether such changes correlated to phylogenetic relationships. For this analysis, we considered all ortholog groups that had at least one DEG responsive to ABA or salt in at least one cell type. There were more DEGs responsive to salt than to ABA in all species (Figure S12). Consequently, the transitions between DEGs and non-DEGs within ortholog groups were more frequent for salt than for ABA (a total of 6,848 and 30,133 ortholog groups for ABA and NaCl response, respectively).

To determine whether the observed patterns of expression divergence occurred above what is expected due to random chance, we developed a bootstrapping approach to generate the expected distribution while preserving the original data structure and computed the *p*-values of observed occurrence for each pattern (see Methods). A gain of response was observed when an ortholog switched its expression from a non-DEG to a DEG, whereas a loss of response was observed when an ortholog switched from a DEG to a non-DEG within one lineage, relative to all other orthologs in the group. Lineage-specific gains of stress responses were consistently observed less than expected and exhibited a wide distribution across cell types and species (Figure 5A). Contrastingly, frequencies for lineage specific loss of stress responses surpassed expected frequencies. These occurrences were confined to pericycle and procambium cells in response to ABA but pervasive across most cell types under salt stress (Figure 5A). This distinctive pattern suggests that transcriptional responses to stress in xylem and phloem cells are particularly stable across species. Synapomorphic and convergent responses were observed at frequencies greater than expected, especially in comparison to divergent responses (Figure 5A).

We identified orthologs exhibiting convergent stress responses shared between the two extremophytes, and found 50 and 91 orthologs responding to ABA and NaCl treatments, respectively (Figure 5A). We also detected 123 and 683 orthologs that were only responsive to ABA and NaCl treatments in non-extremophytes (Figure 5A). While the ABA-responsive orthologs shared between the two non-extremophytes exhibited centered enrichment for nucleic acid metabolism processes, these shared by the extremophytes were mostly associated with response to abiotic stress (Table S7). When exposed to NaCl, defense response pathways were strongly enriched among the orthologs induced only in non-extremphytes and those repressed only in the extrempohytes (Table S7).

Previous studies examining salt stress responses at the organ level in Arabidopsis and its extremophyte relatives have suggested that extremophytes exhibit stress-prepared transcriptomes characterized by minimal overall changes, in contrast to the more pronounced transcriptomic response observed in Arabidopsis (Gong et al., 2005; Lee et al., 2016; Tran et al., 2022). We aimed to investigate whether the stress-prepared state at the organ level in extremophytes resulted from a collective response of individual cells exhibiting a uniform low response or if contrasting responses in different cell types culminated in the absence of a net in the transcriptomic response captured at the organ level. We examined co-expression networks of all orthologs in all cell types across species. This resulted in 18 clusters with distinct co-expression patterns (Figures 5B and S13), most of which were present in all ten cell types, but with differential representations. We then calculated the contribution of each cell type to the overall co-expression pattern observed. We focused on the co-expression clusters that showed constitutive expression in extremophytes under ABA and salt treatments as a stress prepared transcriptomic signal. Within ABA co-expression Cluster 0, where extremophyte orthologs exhibited consistently higher expression before and after ABA exposure, none of the cell-types were significantly contributing more than the others to the resulting constitutive expression (Figure 5B). Conversely, in Cluster 6, characterized by the constitutively low expression in the extremophytes, the predominant contributors were xylem cells (Figure 5B). In response to salt, the high constitutive expression of the extremophytes in Cluster 0 was primarily attributed to lateral root cap and atrichoblast cells, whereas the consistently low expression of the extremophytes in Cluster 6 was predominantly contributed by cortex and pericycle cells (Figure 5B). The co-expressed orthologs from lateral root cap and atrichoblast in Cluster 0 under salt were both predominantly enriched for RNA metabolism- and ribosome biogenesis-related processes (Figure S14), which were also among the most responsive processes in Arabidopsis outer cell layers (Figure 4D). This underscores the tendency for certain cell types to play a more substantial role in shaping the collective stress-prepared signal, as opposed to an equal contribution from all cell types.

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

Camelina sativa is an allohexaploid in the Brassicaceae lineage I that has emerged as an oil crop with a high promise for sustainable biofuel production. *C. sativa* can grow in diverse environments, but its overall growth, seed oil yield, and oil quality are severely impacted by salt and other abiotic stressors (Ahmed et al., 2020; Nadakuduti et al., 2023; Steppuhn et al., 2010). We generated a single cell atlas of 27,441 root cells from *C. sativa* under control, 5 µM ABA, and 100 mM NaCl, as described for the other four species (Figure 6). Using our cross-species cell-type annotation strategy developed for this study, we were able to identify all ten major root cell types in *C. sativa* (Figure 6A), which exhibited a clear separation between inner and outer cell layers as observed for other Brassicaceae species (Figure 6B). The use of individual sub-genomes from *C. sativa* did not significantly affect the integration between *C. sativa* and Arabidopsis or the cell-type assignments (Figure S15), indicating that there was a relatively even representation of transcriptome profiles among the three sub-genomes. This is consistent with the

undifferentiated sub-genome structure previously reported for this species (Blume et al., 2023; Kagale et al., 2014).

Prior work has established that the sub-genome G3 in *C. sativa* exhibited partial expression dominance over the other two sub-genomes (Chaudhary et al., 2020), which was consistent with our findings across cell types and conditions with the exception of columella cells where sub-genome G1 showed a higher mean expression compared to the other sub-genomes under control and ABA-treated conditions (Figure S16). However, it remains unclear whether sub-genomes exhibit biased contributions to the overall stress response. To determine this, we identified DEGs from each sub-genome in each cell type (Table S5). Despite the presence of species specific response, the impact of stress on cell type transcriptomes and the functional processes represented by the all DEGs in *C. sativa* largely resembled these for the other four species (Figure 6C, D and S17). Number of DEGs were not significantly different across all sub-genomes irrespective of whether we considered all detected genes within individual sub-genomes or only the complete homeologs (so called triplets) (Chaudhary et al. 2020) (Figure S18). Consistently, transcriptomic responsiveness measured as AUC scores did not differ significantly among the sub-genomes for most cell types under either treatment (Figure 6E). These together demonstrated that, unlike the expression bias between sub-genomes, stress response did not differentiate the sub-genomes in *C. sativa*, implying the absence of large-scale sub-genome stress response dominance.

The recent polyploidization of the *C sativa* genome (Kagale et al. 2014) allowed us to ask what genomic changes corresponded to divergence in gene expression across the three copies of each gene (the triplet). First, to quantify the coding-sequence divergence, we calculated the nonsynonymous/synonymous rate ratio (ω) for each complete triplet. Second, to assess the regulatory sequence divergences among homeologs, we considered divergence of expression at three different parameters: basal level expression, cell-type preferential expression, and expression changes in response to induced stresses, and computed variations of each parameter for each homeolog. This analysis identified a total of 171, 1384, 115, and 140 candidate triplets exhibiting high divergence (top 1% for each metric) among homeologs in 1) coding sequences, 2) basal expression level, 3) expression specificity, and 4) stress responses, respectively (Figure 6F; Table S8). We noted that these highly divergent triplets tended to be shared between different cell types within each of the defined categories, particularly for basal level expression. However, the overlap among highly divergent triplets across all four defined categories was non-existent, and only 6 out of the 171 fast evolving triplets also showed high expression divergence (Figure 6F). These results indicated that coding-sequence and expression level divergences among triplets are largely uncoupled from one another in *C. sativa*.

DISCUSSION

Comparing stress responses at the cellular level between extremophytes and non-extremophytes holds the potential to identify innovations in stress resilience that may expedite the development of climate-change ready crops. Here, we selected five closely related Brassicaceae species with distinct responses to ABA and NaCl and generated their root single cell profiles. We developed a

two-step approach for cross-species and cross-condition cell-type annotations and confirmed the resulting annotation by *in situ* hybridization of selected markers preferentially expressed in cell types, conditions, or species. In addition to serving as a cell atlas resource for Brassicaceae roots under control and stressed conditions, our method of cell-type identity and response assessments under induced stressed conditions serves as a model for other multi-species and treatment comparisons. Our approach was further applied on an allohexaploid oilseed crop, *C. sativa* to demonstrate the versatile detection and transfer of cell-type response features between diploid and polyploid genomes. We anticipate these data and the methods presented in our study will provide new opportunities and perspectives for comparative single cell transcriptomics to enrich our understanding of cell-type-specific evolution across species. The searchable database we established to disseminate these data will facilitate their broader use by the community.

To date, experimentally validated root cell-type markers in plants have been largely derived from model species like Arabidopsis (Grones et al., 2024; Jin et al., 2022). However, the transferability of these markers to other species where root cell-type markers are unavailable has not been determined. We evaluated the conservation of previously established Arabidopsis cell-type markers in the selected Brassicaceae species and noticed that only 75 of the 241 Arabidopsis markers could serve as cell-type markers across all species tested. The dominant cause for non-functional Arabidopsis markers was the low expression levels among divergent orthologs. This suggested that moderate to high expression of orthologs, in addition to preferential expression in the target cell type, should be an important practical feature of cell-type markers. We have newly identified from our dataset over 300 cell-type preferentially expressed genes that were conserved across these Brassicaceae species (Table S2), similar to the observations among grass species (Guillotin et al., 2023). Contrastingly, a previous study comparing rice and Arabidopsis demonstrated that marker genes for the majority of root cell types are poorly conserved (Liu et al., 2021). This raises interesting questions about the relationship between conservation of cell-type markers and phylogenetic distances among species. For example, at which evolutionary distances does cell-type marker conservation erode and how do genes lose or gain cell-type specificity? Advances in our understanding of cell type evolution will inform innovations in synthetic biology, particularly cell-type specific engineering of traits or metabolite production across species.

Identifying new cell populations across species enabled by scRNA-seq provides an unprecedented opportunity to trace distinct cell states or origins of cell types and infer cell-type evolution (J. Wang et al., 2021). The presence of an extra cortex cell layer specific to *S. irio* and *S. parvula* root anatomy was recapitulated with corresponding scRNA-seq data. A second layer of cortex, called middle cortex, can be also detected in Arabidopsis (Baum et al., 2002; Cui, 2016), but at a later developmental stage of roots. Whether the inner cortex layer in *S. irio* and *S. parvula* is developmentally and functionally similar to the Arabidopsis middle cortex and whether the early development of a second cortex layer presents any evolutionary derived trait remains to be elucidated.

Beyond the comparisons at the stress-neutral condition, our dataset enables interrogation of stress response in each cell type across species. Our results demonstrated that environmental

stresses trigger distinct changes in cell-type transcriptome divergence across species, which typically vary among cell types and stresses, and cell types exhibit differential contributions to stress-responsive co-expression networks across species. These findings highlight the distinct roles of cell types in driving stress response divergence among species and the approaches developed here can therefore facilitate the prioritization of cell types that play more significant roles in determining the divergence.

The cell-type responses of *C. sativa* highlighted transcriptional processes that are shared as well as unique to root cell types in the polyploidy oil crop. The emerging diversity of cell-type responses shows the necessity of examining stress responses of crop species to assess the transferability of response mechanisms from model plants. Although the sub-genome G3 demonstrated a greater contribution of highly expressed genes across the majority of cell types, the three sub-genomes within this allohexaploid had not yet differentiated to the extent of contributing to distinct stress responses in any single cell type. Our findings on homeologs expressing variability at different categories suggest that their evolution largely occurs independently. This adds insights into the fate of homeologs and how we may generate agriculturally preferable neopolyploid crops (Ruiz et al., 2020).

Limitations of current study

We used protoplast-based scRNA-seq to assess stress responses across species at single cell resolution. However, the protoplasting-induced effects have been known as a confounding factor in protoplast-based single cell analyses (Denyer and Timmermans 2022). By selecting single cells over single nuclei preparations, we opted to maximize high data coverage and approximation of the cellular transcriptome (Grones et al. 2024). Given the potential bias introduced at the protoplasting stage of the methods, we have compared gene expression from bulk and pseudo-bulk samples for each species under stress-neutral condition, the results of which were taken into account when interpreting our results. Although we captured most major root cell types in all species tested in the current study, we did not detect cell populations assigned to the quiescent center region as they are far fewer in number compared to other cell types. We did not include cell-type response comparisons when low cell population numbers were detected, like endodermal cells and cells from the vascular tissues in NaCl-treated *C. sativa*. Additionally, we did not have sufficient cell populations representing early development stages of cortex cells to determine the developmental regulators that drive the formation of the additional cortex layer in *S. irio* and *S. parvula*. When roots in some species are larger in size compared to the other species, it would be beneficial to increase the amount of tissues subjected to protoplasting in comparative studies. To capture a thorough understanding of stress responses, it is essential to sample multiple time points (Geng et al., 2013), and conduct tests across varying stress intensities, particularly when comparing extremophytes to non-extremophytes to capture dose-dependent responses. While our study lays the groundwork for understanding stress responses, future research would derive greater value by extending assessments to encompass variations in stress magnitude and duration across multiple time frames. This would enable testing whether extremophytes respond similarly to non-extremophytes, while assessing if their threshold for change significantly differs from that of non-extremophytes. In our study, we

experimentally validated the spatial expression of some selected candidates using HCR. While scRNA-seq and HCR results often exhibit high consistency, we also observed discrepant results between the two techniques. Systematic assessment of this discrepancy in the future will provide guidelines for single cell transcriptome studies.

METHODS

Plant materials and growth conditions

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

Preparation of root samples for scRNA-seq

After seedling growth on plates containing ABA, NaCl, or standard media, primary root tips were cut and placed into a 35 mm petri dish containing a 70 um strainer and 4 mL enzyme solution (1.25% [w/v] Cellulase [“ONOZUKA” R-10, Yakult], 0.1%[w/v] Pectolyase [P-3026, Sigma-Aldrich], 0.4 M Mannitol, 20 mM MES [pH 5.7], 20 mM KCl, 10 mM $CaCl_2$, 0.1% [w/v] bovine serum albumin). The sample dish was rotated at 85 rpm for 50 min at 25°C. The root tips were then transferred to a slide glass and gently pressed with a cover glass using a pencil rubber to release protoplasts. The tissues were transferred back to the enzyme solution and incubated for another 10 minutes. Then, the protoplast solution was filtered through a 70 µm filter, and then through a 40 µm filter twice using 50 mL conical tubes. The filtered solution was centrifuged at 500g for 10 min at 22°C using a swinging bucket rotor and the supernatant was discarded. Protoplasts were resuspended using 500 µL Solution A (0.4 M Mannitol, 20 mM MES [pH 5.7], 20 mM KCl, 10 mM $CaCl_2$, 0.1% [w/v] bovine serum albumin) and transferred to a 2 mL round bottom tube. After centrifugation at 200g for 6 min, the pelleted protoplasts were resuspended with 30–50 µL Solution A to achieve the desired cell concentration (~1,000 protoplasts/µL).

scRNA-seq library construction and sequencing

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

Bulk RNA-seq library preparation and sequencing

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

Orthologous gene sets

We used two different sets of orthologous genes in our study: (1) comparative analyses among Arabidopsis, *E. salsugenium*, *S. irio*, and *S. parvula* were performed using 15,198 unambiguous 1-to-1 ortholog groups between the four species. These 1-to-1 orthologues were identified by comparing the primary protein-coding genes from these species using CLfinder (Oh & Dassanayake, 2019) with MMSeq2 (Steinegger & Söding, 2017) as the aligner with default settings. (2) Species-specific analyses were performed using all genes expressed in the given species. The orthologs between each non-Arabidopsis species (including *E. salsugenium*, *S. irio*, *S. parvula*, and *C. sativa*) and Arabidopsis were identified in following steps: orthologous gene pairs of the protein-coding genes between each non-Arabidopsis species and Arabidopsis were determined reciprocally using MMseqs2 (Steinegger & Söding, 2017) and were subsequently merged to generate a non-redundant list. Only orthologous gene pairs that were considered as orthologs by both OrthoFinder (v2.5.4) (Emms & Kelly, 2019) with granularity -I set to 1.6 and CLfinder were kept. This filtered list was eventually supplemented with reciprocal best hits identified by blastp (Camacho et al. 2009) to generate the final non-redundant orthologues between each non-Arabidopsis species and Arabidopsis.

Raw reads processing

Raw single-cell RNA-seq FASTQ files were analyzed by Cell Ranger (v5.0.1, 10X Genomics) to generate raw gene-by-cell count matrices. For each species, we retrieved the reference genome and corresponding annotation as follows: Arabidopsis (TAIR10) from The Arabidopsis Information Resource (TAIR; https://www.arabidopsis.org/), *C. sativa* (v2) from Brassicaceae Database (BRAD; http://brassicadb.cn), *E. salsugineum* (v1.0) and *S. parvula* (v2.2) from Phytozome 12 (https://phytozome-next.jgi.doe.gov/) (Phytozome ID 574 and 173, respectively),

reference genome for *S. irio* (v1.0) from Brassicaceae Database (BRAD; http://brassicadb.cn) with the updated gene model annotation (Sun et al., 2022) from the CoGe database (https://genomevolution.org/) with the Genome ID 57216. To account for the low accuracy of 3' UTR annotation in non-Arabidopsis species, we extended the 3' UTR of all genes in these species by 500bp, as reported before (Omary et al., 2022). Reads from each species were aligned to the corresponding reference genome and counts were quantified with the Cell Ranger 'count' command with default parameters.

Comparison of gene expression from bulk and pseudo-bulk samples

Raw 150-bp paired-end reads were quality checked with FastQC (v0.11.5; https://www.bioinformatics.babraham.ac.uk/projects/fastqc) and aligned to the respective genomes with HISAT2 (v2.2.1) (Kim et al. 2019). Gene counts were calculated using featureCounts (v2.0.3) (Liao et al. 2014) with the modified gene annotations. Pseudo-bulk gene counts from each scRNA-seq sample were generated by aggregating Cell Ranger-derived counts for all filtered cells. We next implemented edgeR (v3.36.0) (Robinson et al. 2010) to identify genes that exhibited differential expression (FDR ≤ 0.001, absolute $\log_2 FC \geq 2$) between the bulk and pseudo-bulk samples to be aware of any genes that may deviate between how the data was generated. The list is provided in TableS9.

Identification of doublets

Doublets in each scRNA-seq dataset were detected with DoubletFinder (v3) (McGinnis et al., 2019) and subsequently removed prior to any downstream analyses. Three input parameters are required for doublets prediction: the number of expected real doublets (nExp), the number of artificial doublets (pN) and the neighborhood size (pK). We selected a stringent nExp of 7.5% based on the estimated doublet rate in 10X Genomics Chromium Single Cell 3' Reagent Kit User Guide, which was subsequently adjusted according to the proportion of homotypic doublets and doublet ratio. For pN, the default value 25% was used as DoubletFinder performance is largely invariant of pN (McGinnis et al., 2019). The optimal pK value was determined by sequentially loading pre-processed Seurat data into the "paramSweep_v3 (PCs = 1:15)", "summarizeSweep", and "find.pK" functions. For each sample from each species, the single easily discernible maximum pK value was selected as the optimal pK parameter. Only cells which were flagged as singlets were used for downstream analyses.

Data integration, clustering

Data integration and clustering were performed using R package Seurat (v4.0.5) (Stuart et al., 2019). Prior to data integration, we further removed cells in which the numbers of expressed genes were less than 400 and more than 10,000 or < 5% mitochondrial reads, and genes expressed in < 3 cells.

For integration of datasets from each species, we first log-normalized each dataset using the *NormalizeData* function and identified the top 2,000 highly variable genes (HVGs) using a variance-stabilizing transformation in *FindVariableFeatures*. We then used *SelectIntegrationFeatures* to determine 2,000 integration features, which were subsequently used

in *FindIntegrationAnchors* to decide integration anchors between datasets. Datasets from each species were eventually integrated using *IntegrateData* with 30 principal components (PCs). The integrated dataset for each species was scaled and reduced dimensionally with Principal Component Analysis (PCA) followed by Uniform Manifold Approximation and Projection (UMAP) using 50 PCs. *FindNeighbors* and *FindClusters* were used to determine the nearest neighbors and clusters, respectively.

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

Cell type assignment

A) Establishment of a reference atlas using Arabidopsis control samples

Correlation-based annotation: We obtained bulk RNA-seq data (Li et al., 2016) and ATH1 microarray data (Brady et al., 2007) for 14 Arabidopsis root cell types isolated with fluorescence activated cell sorting (FACS) as fragments per kilobase of transcript per million fragments mapped (FPKM) and normalized probe values, respectively. We then computed the expression variance for all genes within each dataset to identify highly variable genes (HVGs) and categorize them into percentile bins. We hypothesized that the transcriptome signature of an individual cell should remain most similar to that of the cell type it originates from regardless of the HVGs used. Therefore, we calculated the Pearson correlation coefficient between each cell from Arabidopsis control samples and the whole-transcriptome reference expression profiles for cell types using HVGs from top 10% to 100% percentile bins. At each percentile bin, each cell was annotated as the cell type with the highest correlation value.

Marker-based annotation: We retrieved the normalized counts and uniform manifold approximation and projection (UMAP) embeddings in Seurat for all cells from Arabidopsis control samples, which were used to calculate the enrichment scores of all genes for each cell using a cluster-independent approach in SEMITONES (Vlot et al., 2022). Similarity between cells was estimated by a radial basis function kernel over the multidimensional UMAP space, and the enrichment scores for all genes were computed using a simple linear regression framework. We subsequently extracted the enrichment scores for 454 (after removing stele markers) known Arabidopsis root cell type marker genes (Table S1), which were compiled through an extensive literature review (Coate et al., 2020; Denyer et al., 2019; Jean-Baptiste et al., 2019; Liu et al., 2021; Ryu et al., 2019; Shahan et al., 2022; Wendrich et al., 2020). To identify statistically significantly enriched markers for individual cells, an empirical distribution of enrichment scores for each cell was obtained by permutating expression data 100 times. A

gene is considered significantly enriched in the given cell if its enrichment score is more than n standard deviations (nsds) away from the mean of the empirical distribution. We determined the significantly enriched markers at nsds of 5, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50 for each cell, and annotated the cell according to the marker with the highest enrichment score at each nsds.

Integration-based annotation: We obtained a previous Arabidopsis root single-cell study (Shahan et al., 2022) with cell type annotations for individual cells. *FindTransferAnchors* from Seurat was used to find transfer anchor cells using the published data as the reference and cells from Arabidopsis control samples in our dataset as the query. *TransferData* was subsequently employed to predict the identity of each cell in the query dataset. These steps were run with different thresholds for the key parameters: ndim at 30, 50 to 500 with an increment interval of 50; kscore at 10, 30, 50; kweight at 10, 20, 30, 50, 80, 100. Cell type prediction for each cell was extracted at every parameter combination.

Combination of annotation methods: We first determined the consensus annotation in each of the three annotation approaches for each cell by identifying the most dominant cell type annotation. The final annotation for each cell in Arabidopsis control samples was assigned by combining the information from the three annotation approaches. If a cell was annotated to the same cell type from at least two of the three approaches, then it was assigned that cell type. In rare cases where a cell cannot be annotated in the aforementioned way, the cell type was assigned according to the marker-based annotation.

B) Annotation transfer from Arabidopsis to other species

The integration of cells from different treatments and different species in our dataset suggested there were no treatment- or species-specific cell populations. We next employed three different reference-based cell type annotation algorithms, including integration-based prediction in Seurat (Stuart et al., 2019), clustering-based prediction in Symphony (Kang et al., 2021), and the machine learning-based prediction in scArches (Lotfollahi et al., 2022), to transfer cell type annotation from Arabidopsis control samples to treatment samples and samples from other four species. To mitigate probabilistic annotations in these methods, we varied the key parameters in each method: Seurat was run with ndim at 30, 50 to 500 with an increment interval of 50, kscore at 10, 30, 50, kweight at 10, 20, 30, 50, 80, 100; Symphony was run with hvg_num from 2000 to 14000 with an interval of 2000, pc_num from 20 to 100 with an interval of 20, and k_num at 5, 10, 20, 30, 50, 80, 100; and scArches was run with hvg_num from 2000 to 14000 with an interval of 2000. Similar to the approach we used to annotate cells in Arabidopsis reference, we next identified the most dominant cell type annotation for a cell as the consensus annotation for that cell in each of the three methods. The most frequent cell type annotation for a cell from all three methods was assigned as the final annotation for that cell. In very rare cases where a cell was annotated differently from the three methods, we compared the average probability given by each method and assigned the cell to the prediction with the highest probability.

Cell-type marker gene identification

To identify cell type marker genes for each species, we used the Seurat *FindMarkers* function on scaled expression using the default Wilcoxon test. Significant markers were selected based on

the following criteria: adjust *p*-value ≤ 0.01, average log fold change ≥ 1, cell type specificity ≥ 0.9 and pct.diff > 0.75, where pct.dff is the ratio of the difference in the percentage of cells expressing the gene between the target cell type and the rest of the cells to the percentage of cells expressing the gene in the target cell type.

Estimation of expression levels and normalization

We estimated the cell-type expression by averaging the expression level (before log transformation) for each gene across all the single cells within each cell type. Then, we calculated the spearman rank correlation matrix of the transcriptomes for the 10 cell types.

Cell type specificity indexes

Cell type specificity index (τ) is based on the Tau metric of tissue specificity (Yanai et al., 2005) as following: $\tau = \frac{\sum_{i=1}^{N}(1-x_i)}{N-1}$, 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.

Identification of orthologues with divergent expression

To identify divergently expressed orthologues (Figure 2B), we compared gene expression of the 15,198 1-to-1 orthologues along the defined cell types across Arabidopsis, *E. salsugenium*, *S. irio*, and *S. parvula*. Within each species, we determined the genes that were differentially expressed between any pairs of cell types with Seurat *FindMarkers* function. We then compiled these genes from all four species and used the normalized cell-type pseudo-bulk expression of these genes as the input to cluster gene expression along the ten cell types using an unsupervised soft clustering method implemented in implemented in the R package Mfuzz (v.2.54.0) (L. Kumar & E Futschik, 2007). The number of clusters and the fuzzification parameter required by Mfuzz were estimated by the Dmin and mestimate functions. Subsequently, we inferred within a phylogenetic framework, as described by Cardoso-Moreira *et al*. (Cardoso-Moreira et al., 2019), the probability that there were changes in the expressions of the 1-to-1 orthologues along the defined cell types, that is, that genes changed their cluster assignment in specific branches, using 0.001 as the probability cut-off. In this framework, the ancestral cluster probabilities along the phylogenetic tree were calculated as the weighted averages from the child-nodes. The weights are given by the inverse branch lengths, which were retrieved from a previous study (Huang et al., 2016) using WebPlotDigitizer (https://automeris.io/WebPlotDigitizer), so that closer child-nodes have more weight.

Cell type prioritization

To estimate the degree of transcriptomic change for each cell type under treatments (Figure 4B), we used Augur (Skinnider et al., 2021) to rank cell types based on their degree of response to biological perturbations in single-cell data. Augur uses a machine-learning frame to quantify the separability of perturbed and unperturbed cells as AUC scores, based on the assumption that cell types more responsive to a perturbation should be more separable than less affected cell types within high-dimensional single-cell measurements. Augur analysis was performed for ABA- and NaCl-treated samples separately compared to the control within each of the five species. The

significance of cell type prioritization for each cell type was determined by comparing the AUC scores from all subsamples of the specific cell type of interest to these of all rest cell types in the given condition pair using one-tailed Wilcoxon rank-sum tests (Andrijevic et al., 2022).

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

Pseudo-bulk differential expression analysis

There exist many statistical methods for differential expression (DE) analysis of single-cell transcriptomics data. However, recent benchmarks have demonstrated that pseudo-bulk methods, which aggregate cell-level counts into sample-level "pseudo-bulk" counts, outperform single-cell DE methods and achieve highest fidelity to the experimental ground truth (Murphy & Skene, 2022; Squair et al., 2021). Therefore, we employed a pseudo-bulk approach to detect genes that responded to the treatments for each cell type in each species. Pseudo-bulk transcriptomes were aggregated for each of these cell populations by summing across raw cell-level counts. Subsequently, the edgeR (v3.36.0) (McCarthy et al., 2012) generalized linear model likelihood ratio test was performed between the treatment and the control. A gene was considered differentially expressed in a given cell population if the false discovery-rate adjusted *p*-value was <= 0.01.

Identification of differential regulations of orthologs between species

We extracted DEGs from each species under each treatment that were included in the 1-to-1 orthologs and compared their response to the treatment across species in a ternary fashion to determine differentially regulated orthologs. To further distinguish observed differential regulation patterns between the orthologs from random sampling (Figure 5A), we started with a fixed order of all 1-to-1 orthologs and then permuted the order of aforementioned DEGs for 10,000 times independently in each species for each cell type and generated an expected occurrence distribution for each pattern when assuming independence between the species. The deviation of the observed occurrence for each pattern was computed as the number of standard deviation away from the mean of the corresponding expected distribution. The *p*-value of the observed occurrence for each pattern was calculated by fitting the observed occurrences into the corresponding bootstrapped distributions using R function *pnorm* and subsequently corrected for multiple testing using the Benjamini-Hochberg procedure in R function *p.adjust*.

Determination of cell type contribution to co-expression pattern

For each cell type under each treatment, orthologs that were considered differentially expressed in at least one of the four species were extracted, and subsequently combined for aggregated clustering. The normalized pseudo-bulk cell type expression levels of these orthologs were

subjected to fuzzy k-means clustering implemented in aerie (Gasch & Eisen, 2002). Ortholog pairs in each of the resulting clusters were further filtered based on: (1) the membership of a given ortholog pair was no less than 0.5; (2) the Pearson correlation between the expression of a given ortholog pair and the mean expression of the cluster it belongs to was no less than 0.8 with a *p*-value smaller than 0.01; (3) the cluster needed to have a minimum of 50 members. To test the representation of each co-expression pattern among cell types under each treatment, we compared the proportion of orthologous DEGs derived from different cell types in each co-expression cluster to that combined from all clusters using R function *phyper* with lower.tail = FALSE. The resulting *p*-values were corrected for multiple testing using the Benjamini-Hochberg procedure in R function *p.adjust*.

Identification of extra Cortex population

To determine potential topological differences in Cortex between species, we performed non-linear dimensional reduction for Cortex cells from each species under control condition using Seurat (Stuart et al., 2019) to generate species-specific UMAP representations. The UMAP representation of Cortex cells in *S. irio* was selected as the reference because of the presence of three distinct subpopulations, which were subsequently annotated as left, middle, and right subpopulation based on Monocle (Qiu et al., 2017; Trapnell et al., 2014) trajectories. Gene that are differentially expressed between the three subpopulations were identified usingSeurat *FindMarkers* function with the default Wilcoxon test.

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

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

To compare the contribution of sub-genomes to the data integration between Arabidopsis and *C. sativa*, we used the latest refined sub-genome structure for *C. sativa* as well as the refined syntenic orthologs between these two species (Chaudhary et al., 2020). The integration was performed separately with all sub-genomes and each sub-genome in *C. sativa* using the default settings in IntegrateData from Seurat. For comparison of basal expression and stress response, we only considered homeologous groups with three complete homeologs. The distribution of DEGs among the three sub-genomes was tested using R function phyper with lower.tail = FALSE. The resulting *p*-values were corrected for multiple testing using the Benjamini-Hochberg procedure in R function p.adjust.

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

We identified fast diverging homeologs at the following levels: 1) coding sequences, 2) basal expression level, 3) expression specificity, and 4) stress responses. Divergence at the coding

level among homeologs was evaluated using the ω value (dN/dS), which was calculated for each complete triplets using CODEML (Álvarez-Carretero et al., 2023), and any homeologous groups where either dN or dS was estimated to be zero were removed from downstream analyses. Divergence in expression and expression specificity were computed as coefficient of variation of variance stabilized expression derived from DESeq2 (Love et al., 2014) and Tau indexes of the homeologs within each homeologous groups. Similarly, we used the standard deviation of the fold changes of the homeologs as the estimation for the divergence in stress responses for each triplet. For each of these categories, only homeologous groups that fall in the top one percentile of the metrics, regardless of cell types, were considered as the fast diverging homeologous groups.

Gene ontology analysis

Enrichment analysis of biological pathways on gene sets of interest was performed using BiNGO (Maere et al., 2005). Gene ontology (GO) annotations for Arabidopsis were retrieved from the GO consortium (http://geneontology.org/) on December 12th, 2021. For non-Arabidopsis species, this analysis was conducted either by replacing gene IDs with their Arabidopsis ortholog IDs (for *E. salsugenium*, *S. irio*, and *S. parvula*) or annotating genes with GO annotations transferred from Arabidopsis (for *C. sativa*). Redundant GO terms with > 50% overlap with similar terms were further clustered using Markov clustering implemented via GOMCL (G. Wang et al., 2020) (https://github.com/Guannan-Wang/GOMCL).

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*. (Oliva et al., 2022) with the following modifications: after hybridization and washes, roots with were left in ClearSee (Kurihara et al., 2015) 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 (Schindelin et al., 2012).

Data and code availability

Data from the current study has been deposited at NCBI under BioProject PRJNA1113801, all processed data can be found at NCBI GEO database under accession code GEO: GSE268881. All original code used throughout the data analysis is publicly available as of the date of publication at the following GitHub repository: https://github.com/Guannan-Wang/BrassicaceaeAtlas.

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.) and the US National Science Foundation (NSF-EDGE NSF-BSF-IOS-EDGE 1923589/2019610 to M.D.). 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. D.G. and J.C. work conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231.

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., J.C., and D.G.; 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.

Declaration of interests

The authors declare no competing interests.

REFERENCES

Ahmed, Z., Liu, J., Waraich, E. A., Yan, Y., Qi, Z., Gui, D., Zeng, F., Tariq, A., Shareef, M., Iqbal, H., Murtaza, G., & Zhang, Z. (2020). Differential physio-biochemical and yield responses of Camelina sativa L. under varying irrigation water regimes in semi-arid climatic conditions. *PloS One*, *15*(12), e0242441.

Álvarez-Carretero, S., Kapli, P., & Yang, Z. (2023). Beginner's Guide on the Use of PAML to

Detect Positive Selection. *Molecular Biology and Evolution*, *40*(4). https://doi.org/10.1093/molbev/msad041

Andrijevic, D., Vrselja, Z., Lysyy, T., Zhang, S., Skarica, M., Spajic, A., Dellal, D., Thorn, S. L., Duckrow, R. B., Ma, S., Duy, P. Q., Isiktas, A. U., Liang, D., Li, M., Kim, S.-K., Daniele, S. G., Banu, K., Perincheri, S., Menon, M. C., … Sestan, N. (2022). Cellular recovery after prolonged warm ischaemia of the whole body. *Nature*, *608*(7922), 405–412.

Baum, S. F., Dubrovsky, J. G., & Rost, T. L. (2002). Apical organization and maturation of the cortex and vascular cylinder inArabidopsis thaliana (Brassicaceae) roots. *American Journal of Botany*, *89*(6), 908–920.

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

Brady, S. M., Orlando, D. A., Lee, J.-Y., Wang, J. Y., Koch, J., Dinneny, J. R., Mace, D., Ohler, U., & Benfey, P. N. (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science*, *318*(5851), 801–806.

Cardoso-Moreira, M., Halbert, J., Valloton, D., Velten, B., Chen, C., Shao, Y., Liechti, A., Ascenção, K., Rummel, C., Ovchinnikova, S., Mazin, P. V., Xenarios, I., Harshman, K., Mort, M., Cooper, D. N., Sandi, C., Soares, M. J., Ferreira, P. G., Afonso, S., … Kaessmann, H. (2019). Gene expression across mammalian organ development. *Nature*, *571*(7766), 505–509.

Chaudhary, R., Koh, C. S., Kagale, S., Tang, L., Wu, S. W., Lv, Z., Mason, A. S., Sharpe, A. G., Diederichsen, A., & Parkin, I. A. P. (2020). Assessing Diversity in the Camelina Genus Provides Insights into the Genome Structure of Camelina sativa. *G3* , *10*(4), 1297–1308.

Choi, H. M. T., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., & Pierce, N. A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* , *145*(12). https://doi.org/10.1242/dev.165753

Coate, J. E., Farmer, A. D., Schiefelbein, J. W., & Doyle, J. J. (2020). Expression Partitioning of Duplicate Genes at Single Cell Resolution in Arabidopsis Roots. *Frontiers in Genetics*, *11*, 596150.

Cui, H. (2016). Middle Cortex Formation in the Root: An Emerging Picture of Integrated Regulatory Mechanisms. *Molecular Plant*, *9*(6), 771–773.

Daryanavard, H., Postiglione, A. E., Mühlemann, J. K., & Muday, G. K. (2023). Flavonols modulate plant development, signaling, and stress responses. *Current Opinion in Plant Biology*, *72*, 102350.

Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., & Timmermans, M. C. P. (2019). Spatiotemporal Developmental Trajectories in the Arabidopsis Root Revealed Using High-Throughput Single-Cell RNA Sequencing. *Developmental Cell*, *48*(6), 840–852.e5.

Dinneny, J. R., Long, T. A., Wang, J. Y., Jung, J. W., Mace, D., Pointer, S., Barron, C., Brady, S. M., Schiefelbein, J., & Benfey, P. N. (2008). Cell identity mediates the response of Arabidopsis roots to abiotic stress. *Science*, *320*(5878), 942–945.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238.

Gasch, A. P., & Eisen, M. B. (2002). Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome Biology*, *3*(11), RESEARCH0059.

Geng, Y., Wu, R., Wee, C. W., Xie, F., Wei, X., Chan, P. M. Y., Tham, C., Duan, L., &

Dinneny, J. R. (2013). A spatio-temporal understanding of growth regulation during the salt stress response in Arabidopsis. *The Plant Cell*, *25*(6), 2132–2154.

Gong, Q., Li, P., Ma, S., Indu Rupassara, S., & Bohnert, H. J. (2005). Salinity stress adaptation competence in the extremophile Thellungiella halophila in comparison with its relative Arabidopsis thaliana. *The Plant Journal: For Cell and Molecular Biology*, *44*(5), 826–839.

Grones, C., Eekhout, T., Shi, D., Neumann, M., Berg, L. S., Ke, Y., Shahan, R., Cox, K. L., Jr, Gomez-Cano, F., Nelissen, H., Lohmann, J. U., Giacomello, S., Martin, O. C., Cole, B., Wang, J.-W., Kaufmann, K., Raissig, M. T., Palfalvi, G., Greb, T., … De Rybel, B. (2024). Best practices for the execution, analysis, and data storage of plant single-cell/nucleus transcriptomics. *The Plant Cell*, *36*(4), 812–828.

Guillotin, B., Rahni, R., Passalacqua, M., Mohammed, M. A., Xu, X., Raju, S. K., Ramírez, C. O., Jackson, D., Groen, S. C., Gillis, J., & Birnbaum, K. D. (2023). A pan-grass transcriptome reveals patterns of cellular divergence in crops. *Nature*, *617*(7962), 785–791.

Huang, C.-H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., Zhang, Q., Koch, M. A., Al-Shehbaz, I., Edger, P. P., Pires, J. C., Tan, D.-Y., Zhong, Y., & Ma, H. (2016). Resolution of Brassicaceae Phylogeny Using Nuclear Genes Uncovers Nested Radiations and Supports Convergent Morphological Evolution. *Molecular Biology and Evolution*, *33*(2), 394–412.

Iyer-Pascuzzi, A. S., Jackson, T., Cui, H., Petricka, J. J., Busch, W., Tsukagoshi, H., & Benfey, P. N. (2011). Cell identity regulators link development and stress responses in the Arabidopsis root. *Developmental Cell*, *21*(4), 770–782.

Jean-Baptiste, K., McFaline-Figueroa, J. L., Alexandre, C. M., Dorrity, M. W., Saunders, L., Bubb, K. L., Trapnell, C., Fields, S., Queitsch, C., & Cuperus, J. T. (2019). Dynamics of Gene Expression in Single Root Cells of Arabidopsis thaliana. *The Plant Cell*, *31*(5), 993–1011.

Jin, J., Lu, P., Xu, Y., Tao, J., Li, Z., Wang, S., Yu, S., Wang, C., Xie, X., Gao, J., Chen, Q., Wang, L., Pu, W., & Cao, P. (2022). PCMDB: a curated and comprehensive resource of plant cell markers. *Nucleic Acids Research*, *50*(D1), D1448–D1455.

Kagale, S., Koh, C., Nixon, J., Bollina, V., Clarke, W. E., Tuteja, R., Spillane, C., Robinson, S. J., Links, M. G., Clarke, C., Higgins, E. E., Huebert, T., Sharpe, A. G., & Parkin, I. A. P. (2014). The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure. *Nature Communications*, *5*, 3706.

Kang, J. B., Nathan, A., Weinand, K., Zhang, F., Millard, N., Rumker, L., Moody, D. B., Korsunsky, I., & Raychaudhuri, S. (2021). Efficient and precise single-cell reference atlas mapping with Symphony. *Nature Communications*, *12*(1), 5890.

Kawa, D., & Brady, S. M. (2022). Root cell types as an interface for biotic interactions. *Trends in Plant Science*, *27*(11), 1173–1186.

Kazachkova, Y., Eshel, G., Pantha, P., Cheeseman, J. M., Dassanayake, M., & Barak, S. (2018). Halophytism: What Have We Learnt From Arabidopsis thaliana Relative Model Systems? *Plant Physiology*, *178*(3), 972–988.

Kumar, L., & E Futschik, M. (2007). Mfuzz: a software package for soft clustering of microarray data. *Bioinformation*, *2*(1), 5–7.

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

Kurihara, D., Mizuta, Y., Sato, Y., & Higashiyama, T. (2015). ClearSee: a rapid optical clearing reagent for whole-plant fluorescence imaging. *Development* , *142*(23), 4168–4179.

Lee, Y. P., Funk, C., Erban, A., Kopka, J., Köhl, K. I., Zuther, E., & Hincha, D. K. (2016). Salt

stress responses in a geographically diverse collection of Eutrema/Thellungiella spp. accessions. *Functional Plant Biology: FPB*, *43*(7), 590–606.

Li, S., Yamada, M., Han, X., Ohler, U., & Benfey, P. N. (2016). High-Resolution Expression Map of the Arabidopsis Root Reveals Alternative Splicing and lincRNA Regulation. *Developmental Cell*, *39*(4), 508–522.

Liu, Q., Liang, Z., Feng, D., Jiang, S., Wang, Y., Du, Z., Li, R., Hu, G., Zhang, P., Ma, Y., Lohmann, J. U., & Gu, X. (2021). Transcriptional landscape of rice roots at the single-cell resolution. *Molecular Plant*, *14*(3), 384–394.

Lotfollahi, M., Naghipourfar, M., Luecken, M. D., Khajavi, M., Büttner, M., Wagenstetter, M., Avsec, Ž., Gayoso, A., Yosef, N., Interlandi, M., Rybakov, S., Misharin, A. V., & Theis, F. J. (2022). Mapping single-cell data to reference atlases by transfer learning. *Nature Biotechnology*, *40*(1), 121–130.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550.

Maere, S., Heymans, K., & Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* , *21*(16), 3448–3449.

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, *40*(10), 4288–4297.

McGinnis, C. S., Murrow, L. M., & Gartner, Z. J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Systems*, *8*(4), 329–337.e4.

Murphy, A. E., & Skene, N. G. (2022). A balanced measure shows superior performance of pseudobulk methods in single-cell RNA-sequencing analysis [Review of *A balanced measure shows superior performance of pseudobulk methods in single-cell RNA-sequencing analysis*]. *Nature Communications*, *13*(1), 7851.

Nadakuduti, S. S., Laforest, L. C., Tachev, M., Decker, A. N., Egesa, A. O., Shirazi, A. S., Begcy, K., Sarnoski, P. J., & Buell, C. R. (2023). Heat stress during seed development leads to impaired physiological function and plasticity in seed oil accumulation in Camelina sativa. *Frontiers in Plant Science*, *14*, 1284573.

Oh, D.-H., & Dassanayake, M. (2019). Landscape of gene transposition-duplication within the Brassicaceae family. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *26*(1), 21–36.

Oh, D.-H., Hong, H., Lee, S. Y., Yun, D.-J., Bohnert, H. J., & Dassanayake, M. (2014). Genome structures and transcriptomes signify niche adaptation for the multiple-ion-tolerant extremophyte Schrenkiella parvula. *Plant Physiology*, *164*(4), 2123–2138.

Oliva, M., Stuart, T., Tang, D., Pflueger, J., Poppe, D., Jabbari, J. S., Gigante, S., Dragwidge, J. M., Whelan, J., Lewsey, M. G., & Lister, R. (2022). An environmentally responsive transcriptional state modulates cell identities during root development. In *bioRxiv* (p. 2022.03.04.483008). https://doi.org/10.1101/2022.03.04.483008

Omary, M., Gil-Yarom, N., Yahav, C., Steiner, E., Hendelman, A., & Efroni, I. (2022). A conserved superlocus regulates above- and belowground root initiation. *Science*, *375*(6584), eabf4368.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., & Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nature Methods*,

14(10), 979–982.
Ruiz, M., Oustric, J., Santini, J., & Morillon, R. (2020). Synthetic Polyploidy in Grafted Crops. *Frontiers in Plant Science*, *11*, 540894.
Ryu, K. H., Huang, L., Kang, H. M., & Schiefelbein, J. (2019). Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. *Plant Physiology*, *179*(4), 1444–1456.
Sappl, P. G., Carroll, A. J., Clifton, R., Lister, R., Whelan, J., Harvey Millar, A., & Singh, K. B. (2009). The Arabidopsis glutathione transferase gene family displays complex stress regulation and co-silencing multiple genes results in altered metabolic sensitivity to oxidative stress. *The Plant Journal: For Cell and Molecular Biology*, *58*(1), 53–68.
Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682.
Shahan, R., Hsu, C.-W., Nolan, T. M., Cole, B. J., Taylor, I. W., Greenstreet, L., Zhang, S., Afanassiev, A., Vlot, A. H. C., Schiebinger, G., Benfey, P. N., & Ohler, U. (2022). A single-cell Arabidopsis root atlas reveals developmental trajectories in wild-type and cell identity mutants. *Developmental Cell*, *57*(4), 543–560.e9.
Skinnider, M. A., Squair, J. W., Kathe, C., Anderson, M. A., Gautier, M., Matson, K. J. E., Milano, M., Hutson, T. H., Barraud, Q., Phillips, A. A., Foster, L. J., La Manno, G., Levine, A. J., & Courtine, G. (2021). Cell type prioritization in single-cell data. *Nature Biotechnology*, *39*(1), 30–34.
Squair, J. W., Gautier, M., Kathe, C., Anderson, M. A., James, N. D., Hutson, T. H., Hudelle, R., Qaiser, T., Matson, K. J. E., Barraud, Q., Levine, A. J., La Manno, G., Skinnider, M. A., & Courtine, G. (2021). Confronting false discoveries in single-cell differential expression. *Nature Communications*, *12*(1), 5692.
Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, *35*(11), 1026–1028.
Steppuhn, H., Falk, K. C., & Zhou, R. (2010). Emergence, height, grain yield and oil content of camelina and canola grown in saline media. *Canadian Journal of Soil Science*, *90*(1), 151–164.
Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, *177*(7), 1888–1902.e21.
Sun, Y., Oh, D.-H., Duan, L., Ramachandran, P., Ramirez, A., Bartlett, A., Tran, K.-N., Wang, G., Dassanayake, M., & Dinneny, J. R. (2022). Divergence in the ABA gene regulatory network underlies differential growth control. *Nature Plants*, *8*(5), 549–560.
Taji, T., Seki, M., Satou, M., Sakurai, T., Kobayashi, M., Ishiyama, K., Narusaka, Y., Narusaka, M., Zhu, J.-K., & Shinozaki, K. (2004). Comparative genomics in salt tolerance between Arabidopsis and aRabidopsis-related halophyte salt cress using Arabidopsis microarray. *Plant Physiology*, *135*(3), 1697–1709.
Takano, J., Wada, M., Ludewig, U., Schaaf, G., von Wirén, N., & Fujiwara, T. (2006). The Arabidopsis major intrinsic protein NIP5;1 is essential for efficient boron uptake and plant development under boron limitation. *The Plant Cell*, *18*(6), 1498–1509.
Terrón-Camero, L. C., Peláez-Vico, M. Á., Rodríguez-González, A., Del Val, C., Sandalio, L. M., & Romero-Puertas, M. C. (2022). Gene network downstream plant stress response

modulated by peroxisomal H2O2. *Frontiers in Plant Science*, *13*, 930721.
Tran, K.-N., Wang, G., Oh, D.-H., Larkin, J. C., Smith, A. P., & Dassanayake, M. (2022). Multiple paths lead to salt tolerance - pre-adaptation vs dynamic responses from two closely related extremophytes. In *bioRxiv* (p. 2021.10.23.465591). https://doi.org/10.1101/2021.10.23.465591
Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S., & Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, *32*(4), 381–386.
Vlot, A. H. C., Maghsudi, S., & Ohler, U. (2022). Cluster-independent marker feature identification from single-cell omics data using SEMITONES. *Nucleic Acids Research*, *50*(18), e107.
Waadt, R., Seller, C. A., Hsu, P.-K., Takahashi, Y., Munemasa, S., & Schroeder, J. I. (2022). Plant hormone regulation of abiotic stress responses. *Nature Reviews. Molecular Cell Biology*, *23*(10), 680–694.
Wang, G., DiTusa, S. F., Oh, D.-H., Herrmann, A. D., Mendoza-Cozatl, D. G., O'Neill, M. A., Smith, A. P., & Dassanayake, M. (2021). Cross species multi-omics reveals cell wall sequestration and elevated global transcript abundance as mechanisms of boron tolerance in plants. *The New Phytologist*, *230*(5), 1985–2000.
Wang, G., Oh, D.-H., & Dassanayake, M. (2020). GOMCL: a toolkit to cluster, evaluate, and extract non-redundant associations of Gene Ontology-based functions. *BMC Bioinformatics*, *21*(1), 139.
Wang, J., Sun, H., Jiang, M., Li, J., Zhang, P., Chen, H., Mei, Y., Fei, L., Lai, S., Han, X., Song, X., Xu, S., Chen, M., Ouyang, H., Zhang, D., Yuan, G.-C., & Guo, G. (2021). Tracing cell-type evolution by cross-species comparison of cell atlases. *Cell Reports*, *34*(9), 108803.
Wendrich, J. R., Yang, B., Vandamme, N., Verstaen, K., Smet, W., Van de Velde, C., Minne, M., Wybouw, B., Mor, E., Arents, H. E., Nolf, J., Van Duyse, J., Van Isterdael, G., Maere, S., Saeys, Y., & De Rybel, B. (2020). Vascular transcription factors guide plant epidermal responses to limiting phosphate conditions. *Science*, *370*(6518). https://doi.org/10.1126/science.aay4970
Xu, G., Guo, C., Shan, H., & Kong, H. (2012). Divergence of duplicate genes in exon-intron structure. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(4), 1187–1192.
Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., Lancet, D., & Shmueli, O. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* , *21*(5), 650–659.
Zhang, H., Zhao, Y., & Zhu, J.-K. (2020). Thriving under Stress: How Plants Balance Growth and the Stress Response. *Developmental Cell*, *55*(5), 529–543.

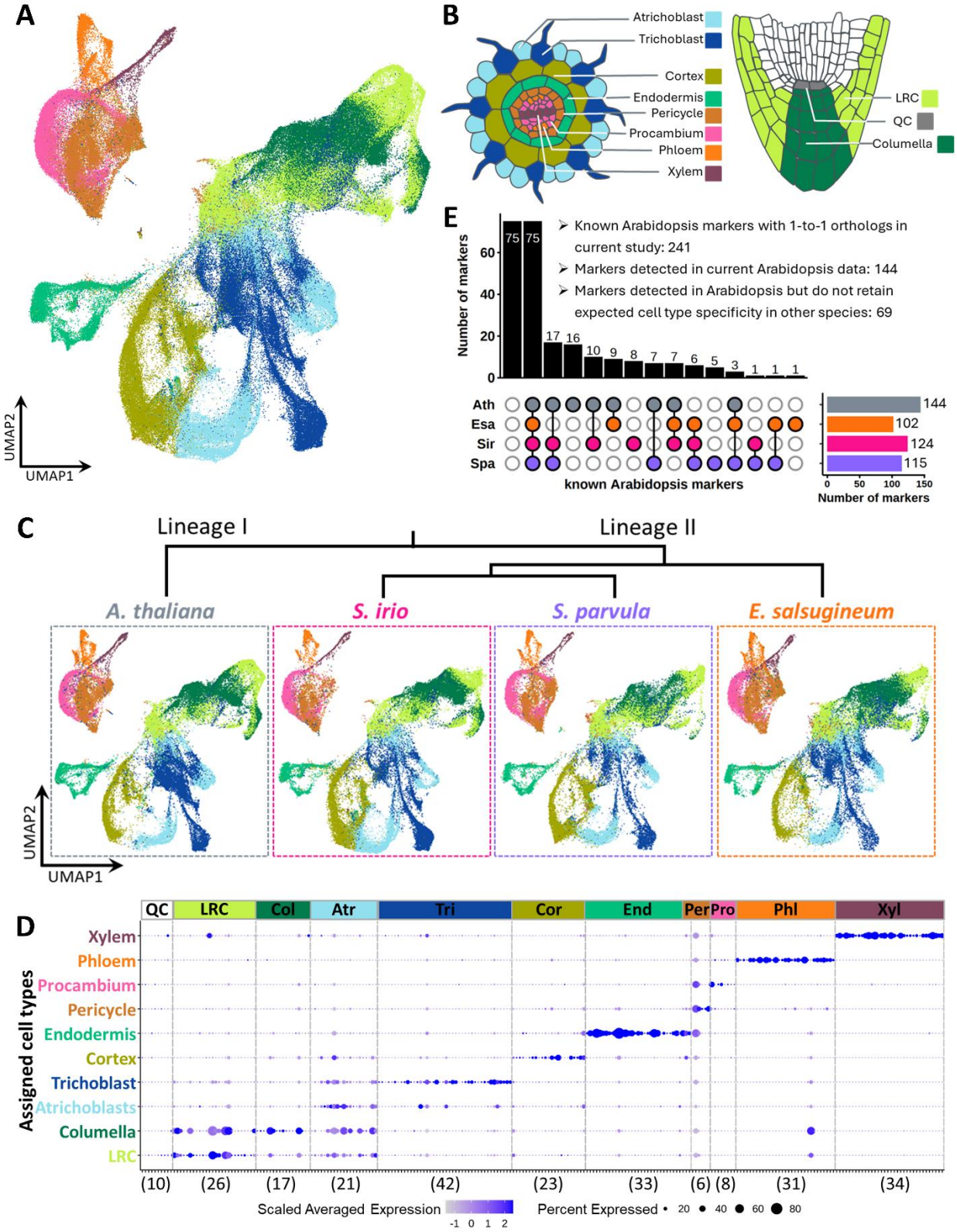

A
UMAP2
UMAP1
B
Atrichoblast
Trichoblast
Cortex
Endodermis
Pericycle
Procambium
Phloem
Xylem
LRC
QC
Columella
E
Number of markers
75 75 17 16 10 9 8 7 7 6 5 3 1 1 1
Known Arabidopsis markers with 1-to-1 orthologs in current study: 241
Markers detected in current Arabidopsis data: 144
Markers detected in Arabidopsis but do not retain expected cell type specificity in other species: 69
Ath
Esa
Sir
Spa
known Arabidopsis markers
144
102
124
115
Number of markers
C
Lineage I
Lineage II
A. thaliana
S. irio
S. parvula
E. salsugineum
UMAP2
UMAP1
D
QC
LRC
Col
Atr
Tri
Cor
End
Per
Pro
Phl
Xyl
Assigned cell types
Xylem
Phloem
Procambium
Pericycle
Endodermis
Cortex
Trichoblast
Atrichoblasts
Columella
LRC
(10) (26) (17) (21) (42) (23) (33) (6) (8) (31) (34)
Scaled Averaged Expression
-1 0 1 2
Percent Expressed
20 40 60 80

Figure 1 Single cell RNA profiling of root cells across 4 Brassicaceae species: Arabidopsis, *Sisymbrium irio*, *Schrenkiella parvula*, and *Eutrema salsugineum*. (A) Integrated UMAP view of root atlas (174,820 cells) from 4 Brassicaceae species representing 10 major root cell types. (B) Schematic illustration of major Arabidopsis root cell types annotated in current study. Illustration adapted from the Plant Illustrations repository (Bouché, 2017) (C) Root cell atlas for individual species on the integrated UMAP space. (D) Preferential expression of known cell-type markers based on cells from all species. See Figure S4 for detailed preferential marker gene expression in each of the four species. See Table S1 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. (E) Number of known Arabidopsis markers that were detected in each species in the current study. QC, Quiescent center; Col, Columella; LRC, Lateral Root Cap; Atr, Atrichoblast; Tri, Trichoblast; Cor, Cortex; End, Endodermis; Per, Pericycle; Pro, Procambium; Phl, Phloem; Xyl, Xylem.

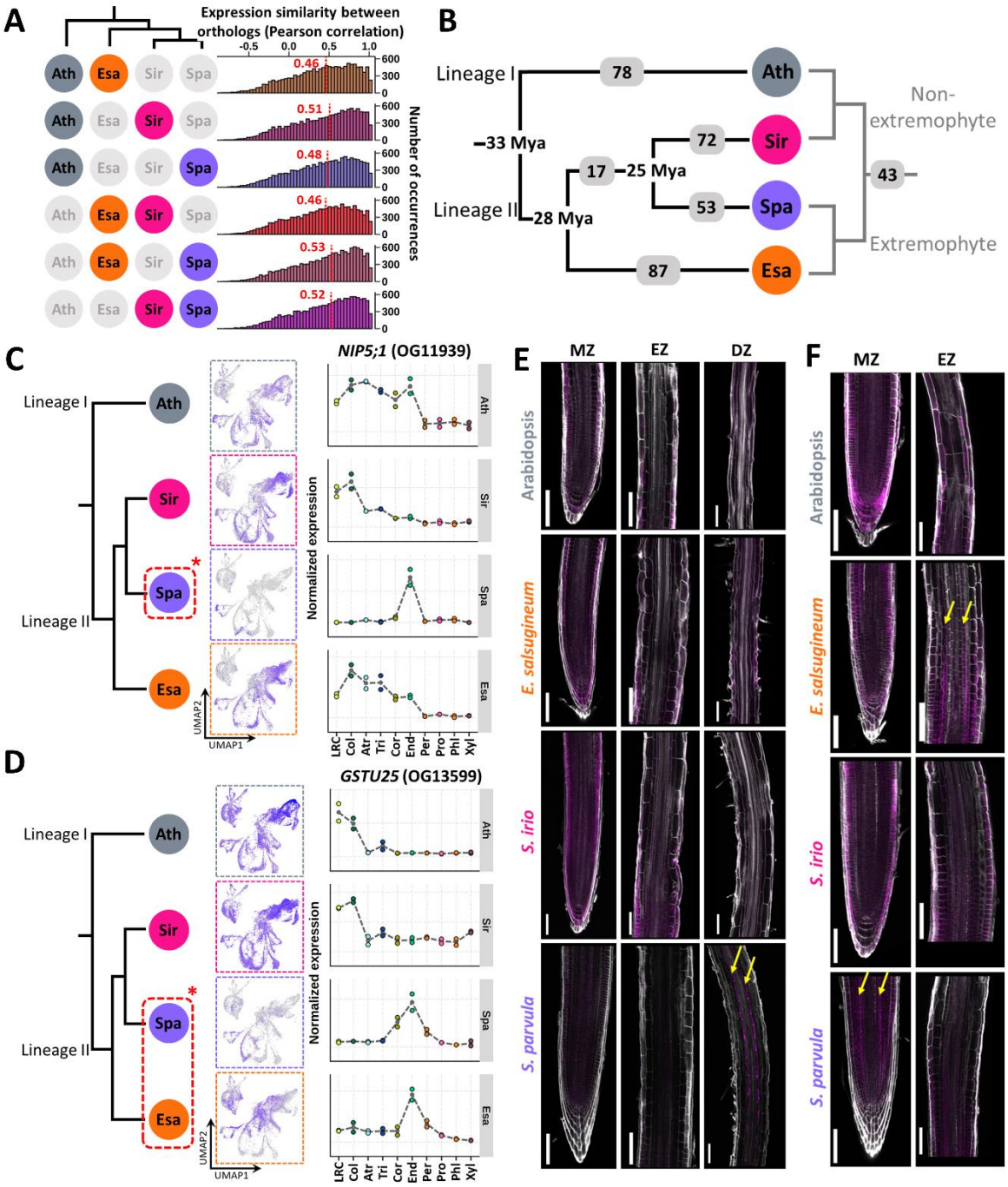

A
Expression similarity between orthologs (Pearson correlation)
-0.5
0.0
0.5
1.0
Ath
Esa
Sir
Spa
0.46
0.51
0.48
0.46
0.53
0.52
600
300
0
Number of occurrences
B
Lineage I
Lineage II
33 Mya
28 Mya
25 Mya
78
17
72
53
87
43
Ath
Sir
Spa
Esa
Non-extremophyte
Extremophyte
C
NIP5;1 (OG11939)
Lineage I
Lineage II
Normalized expression
UMAP1
UMAP2
LRC
Col
Atr
Tri
Cor
End
Per
Pro
Phl
Xyl
D
GSTU25 (OG13599)
E
MZ
EZ
DZ
Arabidopsis
E. salsugineum
S. irio
S. parvula
F
MZ
EZ

Figure 2 Evolution of gene expression profiles across cell types. (A) Distribution of Pearson correlation coefficients for the cell type expression of orthologs between each pair of species. The line and text in red indicate the median. (B) The numbers of orthologs showing diverged expression profiles between species groups. (C, D) Cell type expression of selected examples with diverged expression across species, *NIP5;1* (C) and *GSTU25* (D). 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. Col, Columella; LRC, Lateral Root Cap; 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). Arrows in yellow mark the cell type where the diverged expression is expected. MZ, meristem zone; EZ, elongation zone; DZ, differentiation zone. Scale bar, 100 μm.

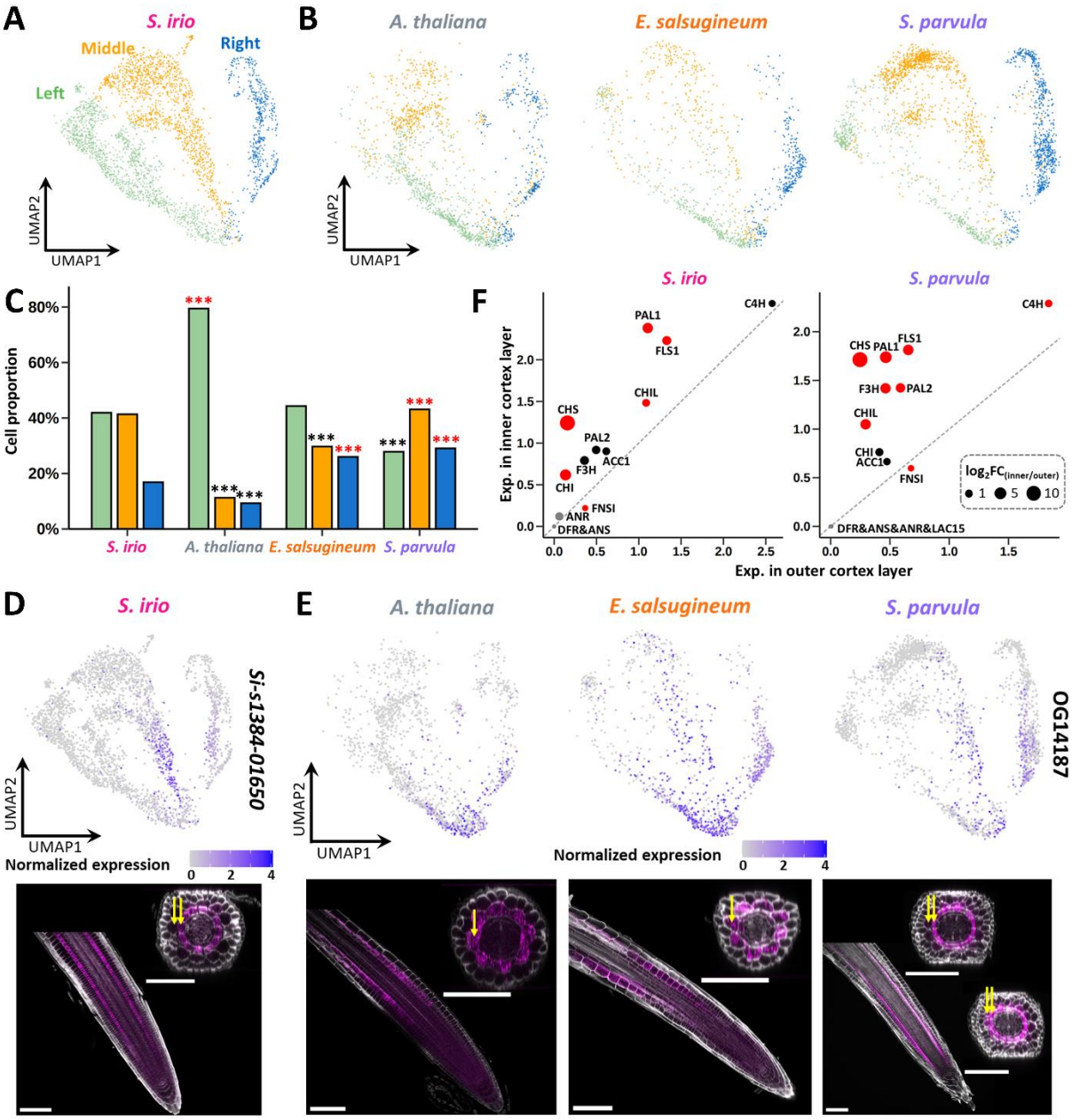

A
S. irio
Middle
Right
Left
UMAP2
UMAP1
B
A. thaliana
E. salsugineum
S. parvula
C
Cell proportion
80%
60%
40%
20%
0%
***
S. irio
A. thaliana
E. salsugineum
S. parvula
F
S. irio
S. parvula
Exp. in inner cortex layer
Exp. in outer cortex layer
C4H
PAL1
FLS1
CHIL
CHS
PAL2
ACC1
F3H
CHI
FNSI
ANR
DFR&ANS
CHS
PAL1
FLS1
F3H
PAL2
CHIL
CHI
ACC1
FNSI
DFR&ANS&ANR&LAC15
log2FC(inner/outer)
1
5
10
D
S. irio
Si-s1384-01650
UMAP2
UMAP1
Normalized expression
0
2
4
E
A. thaliana
E. salsugineum
S. parvula
OG14187
UMAP2
UMAP1
Normalized expression
0
2
4

Figure 3 Comparison of cortex subpopulations among species. (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 (p-value ≤ 0.01) with red asterisks indicating over-representation and black asterisks indicating under-representation. (D) Normalized expression of *Si-s1384-01650* 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 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.

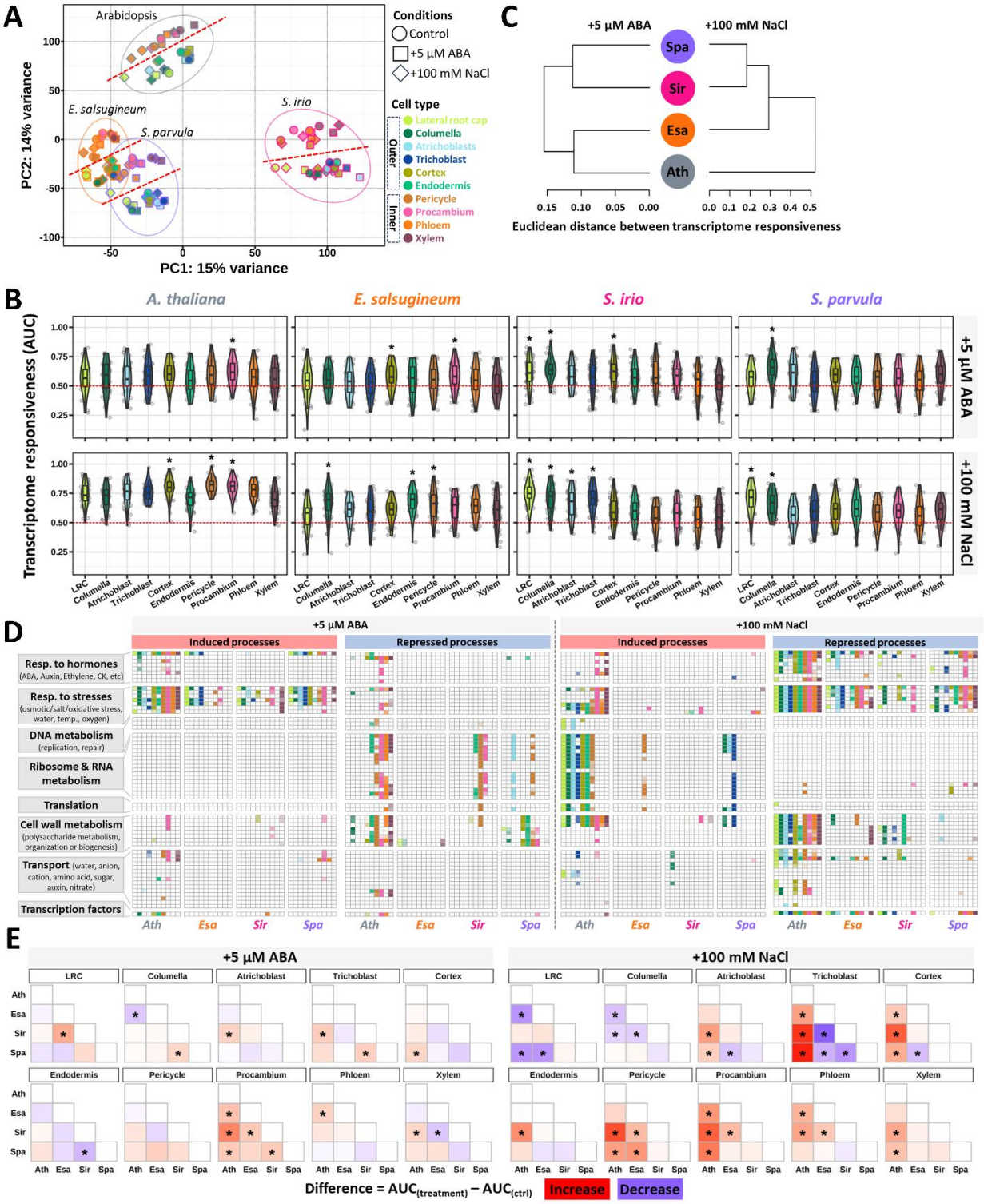

A
Arabidopsis
E. salsugineum
S. parvula
S. irio
PC2: 14% variance
PC1: 15% variance
Conditions
Control
+5 µM ABA
+100 mM NaCl
Cell type
Lateral root cap
Columella
Atrichoblasts
Trichoblast
Cortex
Endodermis
Pericycle
Procambium
Phloem
Xylem
Outer
Inner
C
+5 µM ABA
+100 mM NaCl
Spa
Sir
Esa
Ath
Euclidean distance between transcriptome responsiveness
B
A. thaliana
E. salsugineum
S. irio
S. parvula
Transcriptome responsiveness (AUC)
+5 µM ABA
+100 mM NaCl
LRC
Columella
Atrichoblast
Trichoblast
Cortex
Endodermis
Pericycle
Procambium
Phloem
Xylem
D
+5 µM ABA
+100 mM NaCl
Induced processes
Repressed processes
Resp. to hormones (ABA, Auxin, Ethylene, CK, etc)
Resp. to stresses (osmotic/salt/oxidative stress, water, temp., oxygen)
DNA metabolism (replication, repair)
Ribosome & RNA metabolism
Translation
Cell wall metabolism (polysaccharide metabolism, organization or biogenesis)
Transport (water, anion, cation, amino acid, sugar, auxin, nitrate)
Transcription factors
Ath
Esa
Sir
Spa
E
+5 µM ABA
+100 mM NaCl
LRC
Columella
Atrichoblast
Trichoblast
Cortex
Endodermis
Pericycle
Procambium
Phloem
Xylem
Difference = AUC(treatment) – AUC(ctrl)
Increase
Decrease

Figure 4 Divergent cell type transcriptomic response to ABA and NaCl. (A) Principal component (PC) analysis of pseudo-bulk cell-type transcriptomes. 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 labelled. 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. Asterisks indicate significance after Wilcox testing 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 Table S6. The shade represents the *p*-value assigned by the enrichment test (adjusted *p*-value < 0.05) with darker shades indicating smaller *p*-values. (E) Changes in cell type transcriptome divergence between species in response to ABA (left) and NaCl (right). The darker colors indicate larger differences in transcriptome divergence induced by the treatment. Asterisks indicate significance after comparing to control with Student's t-test at *p*-value cutoff of 0.05. Ath, Arabidopsis; Esa, *Eutrema salsugineum*; Sir, *Sisymbrium irio;* Spa, *Schrenkiella parvula*.

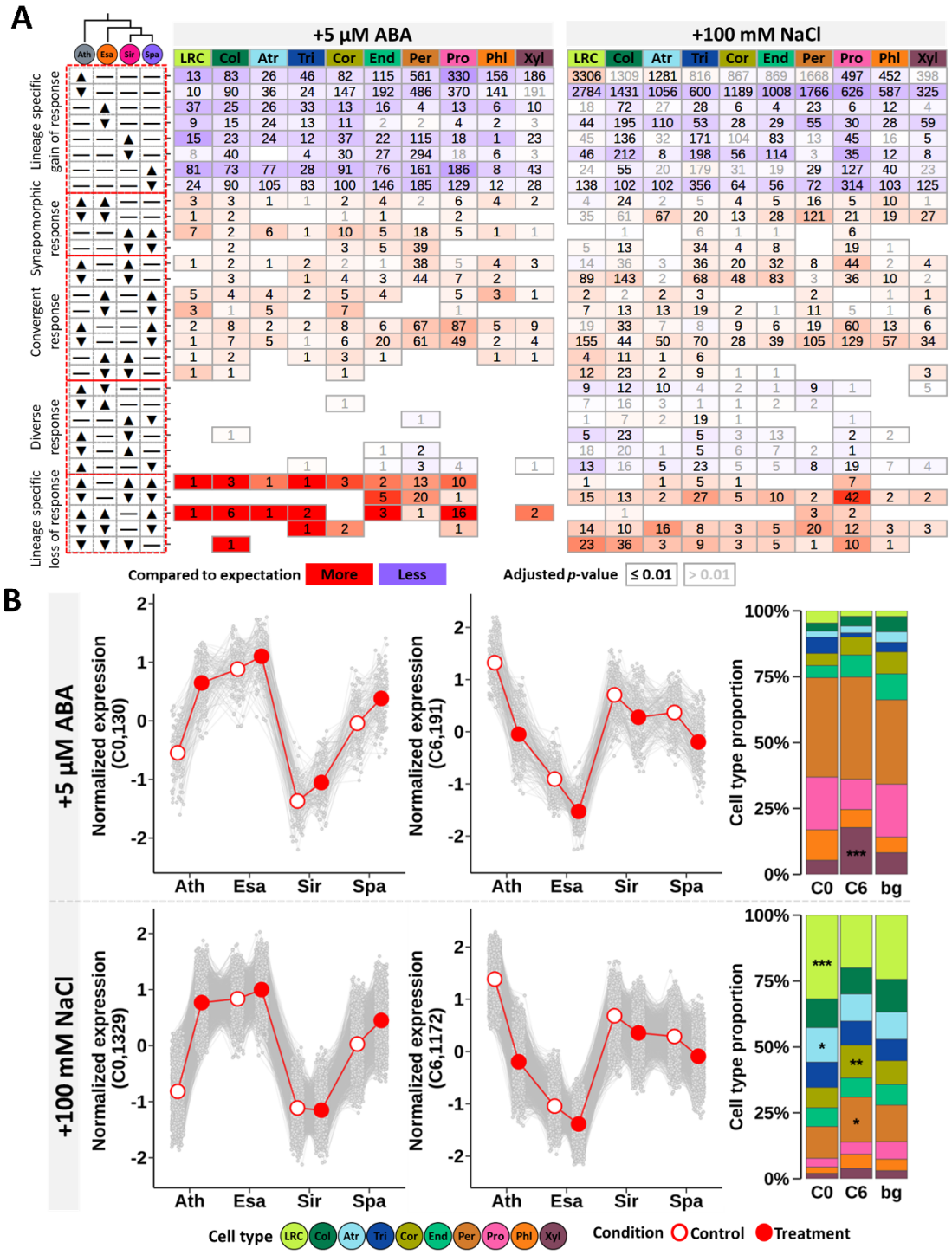

A
Ath
Esa
Sir
Spa
Lineage specific gain of response
Synapomorphic response
Convergent response
Diverse response
Lineage specific loss of response
+5 µM ABA
+100 mM NaCl
LRC Col Atr Tri Cor End Per Pro Phl Xyl
Compared to expectation
More
Less
Adjusted p-value
≤ 0.01
> 0.01
B
+5 µM ABA
+100 mM NaCl
Normalized expression (C0,130)
Normalized expression (C6,191)
Normalized expression (C0,1329)
Normalized expression (C6,1172)
Cell type proportion
100%
75%
50%
25%
0%
C0
C6
bg
Cell type
Condition
Control
Treatment

Figure 5 Stress responses between orthologs. (A) Differences between observed and bootstrapped (10,000 times) occurrences of different ortholog response patterns. Number of orthologs for each pattern in each cell types are indicated in the corresponding tile. The differences were measured as the number of standard deviation 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 the observed occurrences deviate from the expected distribution significantly or not significantly, respectively, at Benjamini-Hochberg-adjusted p-value ≤ 0.01. (B) 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. Col, Columella; LRC, Lateral Root Cap; 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*.

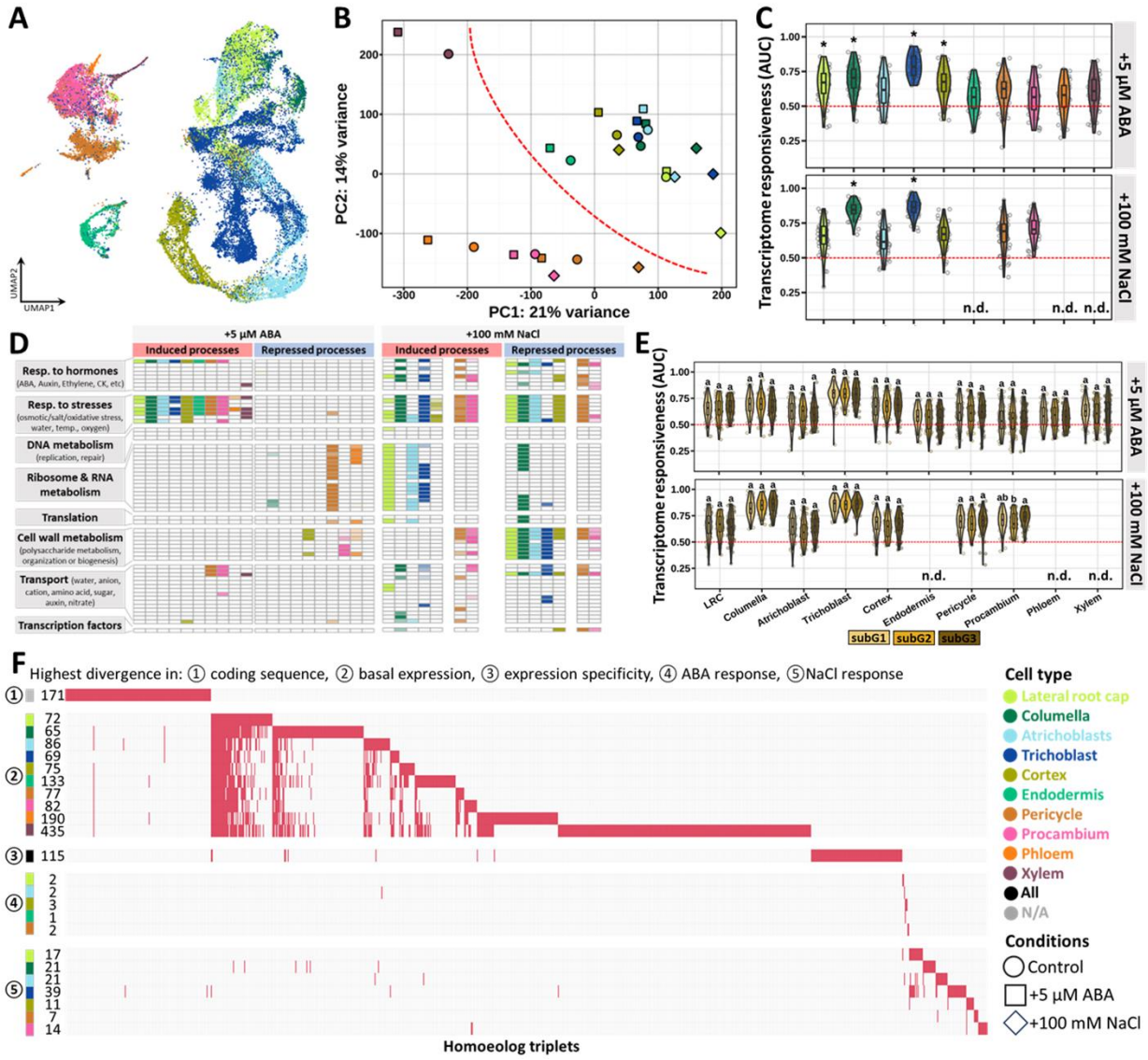
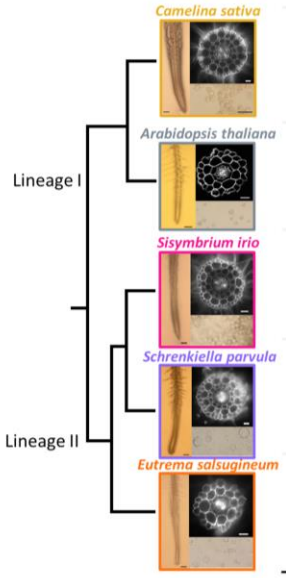

A
UMAP2
UMAP1
B
PC2: 14% variance
PC1: 21% variance
C
Transcriptome responsiveness (AUC)
+5 µM ABA
+100 mM NaCl
n.d.
D
+5 µM ABA
+100 mM NaCl
Induced processes
Repressed processes
Resp. to hormones (ABA, Auxin, Ethylene, CK, etc)
Resp. to stresses (osmotic/salt/oxidative stress, water, temp., oxygen)
DNA metabolism (replication, repair)
Ribosome & RNA metabolism
Translation
Cell wall metabolism (polysaccharide metabolism, organization or biogenesis)
Transport (water, anion, cation, amino acid, sugar, auxin, nitrate)
Transcription factors
E
LRC
Columella
Atrichoblast
Trichoblast
Cortex
Endodermis
Pericycle
Procambium
Phloem
Xylem
subG1
subG2
subG3
F
Highest divergence in: ① coding sequence, ② basal expression, ③ expression specificity, ④ ABA response, ⑤NaCl response
Homoeolog triplets
Cell type
Lateral root cap
Columella
Atrichoblasts
Trichoblast
Cortex
Endodermis
Pericycle
Procambium
Phloem
Xylem
All
N/A
Conditions
Control
+5 µM ABA
+100 mM NaCl

Figure 6 Cell type stress responses in *C. sativa*. (A) Root single cell atlas (27,442 cells) for *C. sativa*. (B) Principal component analysis of pseudo-bulk cell-type transcriptomes. 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. (C) Comparison of transcriptome responsiveness as averaged Augur area under the curve (AUC) scores across cell types in response to ABA and NaCl. Asterisks indicate significance after Wilcox testing 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 Table S6. The shade represents the *p*-value assigned by the enrichment test (false discovery rate (FDR)-adjusted *p*-value < 0.05) with darker shades indicating smaller *p*-values. (E) Comparison of transcriptome responsiveness for each sub-genome in *C. sativa*. Lowercase letters on top of violin plots indicate significantly different groups as determined by one-way analysis of variance (ANOVA) with post-hoc Tukey honest significant difference (HSD) testing (*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	
Camelina sativa	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
Arabidopsis thaliana	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
Sisymbrium irio	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
Schrenkiella parvula	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
Eutrema salsugineum	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

Figure S1 Overview of scRNA-seq dataset for four 5 Brassicaceae species and data preprocessing.

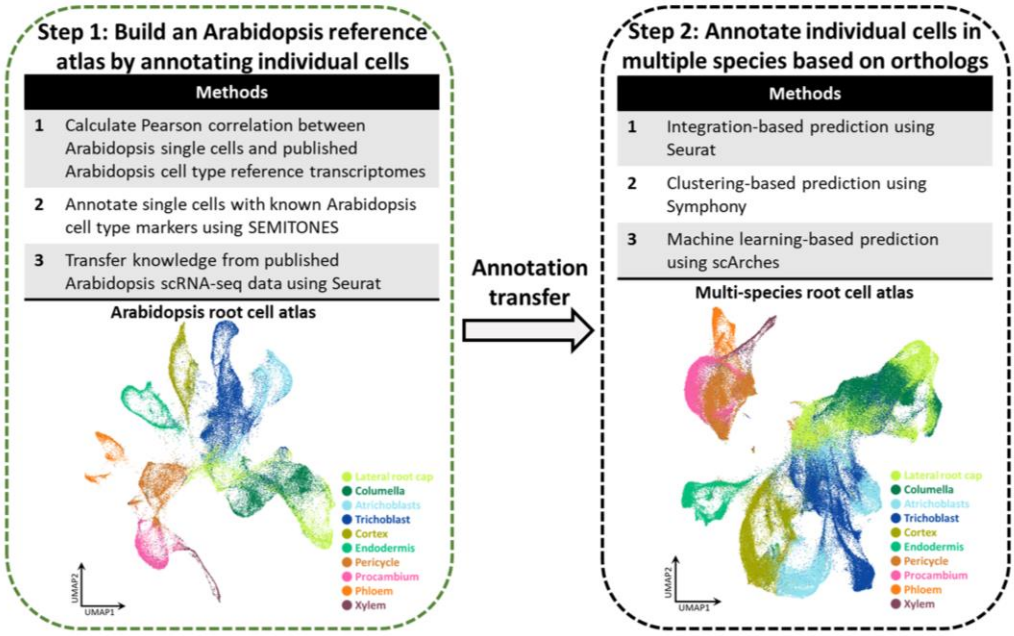


Figure S2 A two-step strategy for cross-species cell type annotations.

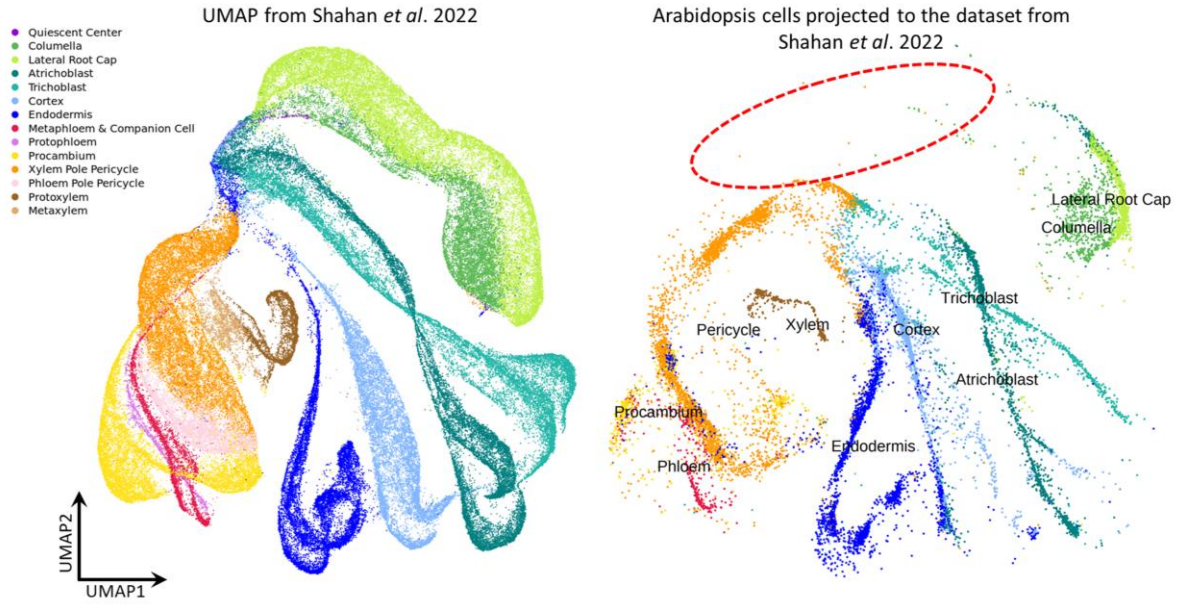


Figure S3 Projection of Arabidopsis control cells from current study to Shahan *et al*.

Figure S4 Cell-type expression of known Arabidopsis cell-type markers and their 1-to-1 ortholog in other species. See Table S1 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; Col, Columella; LRC, Lateral Root Cap; Atr, Atrichoblast; Tri, Trichoblast; Cor, Cortex; End, Endodermis; Per, Pericycle; Pro, Procambium; Phl, Phloem; Xyl, Xylem.

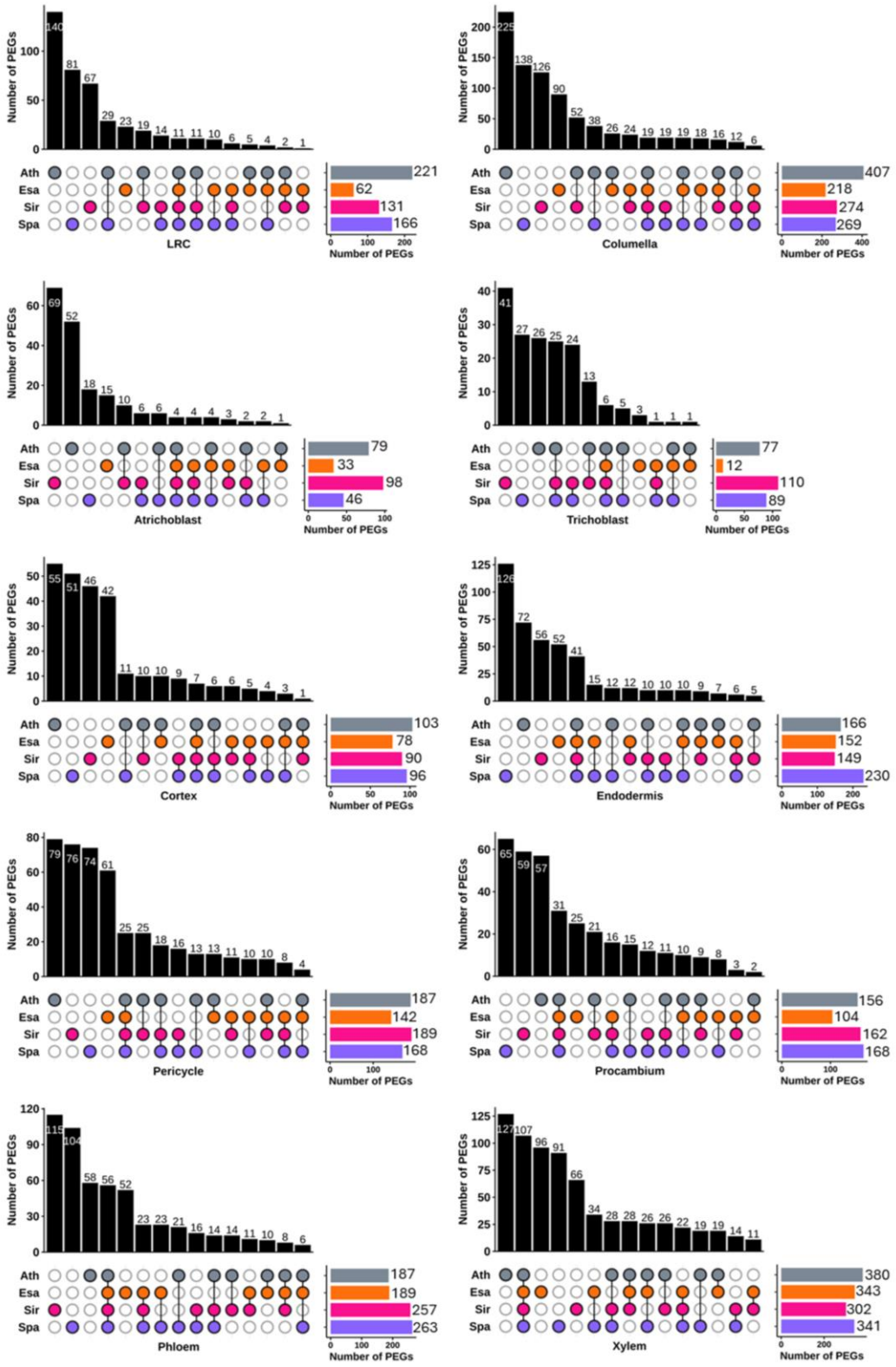


Figure S5 Identification of cell type preferentially enriched genes (PEGs) that were conserved across species for each cell type in the current study.

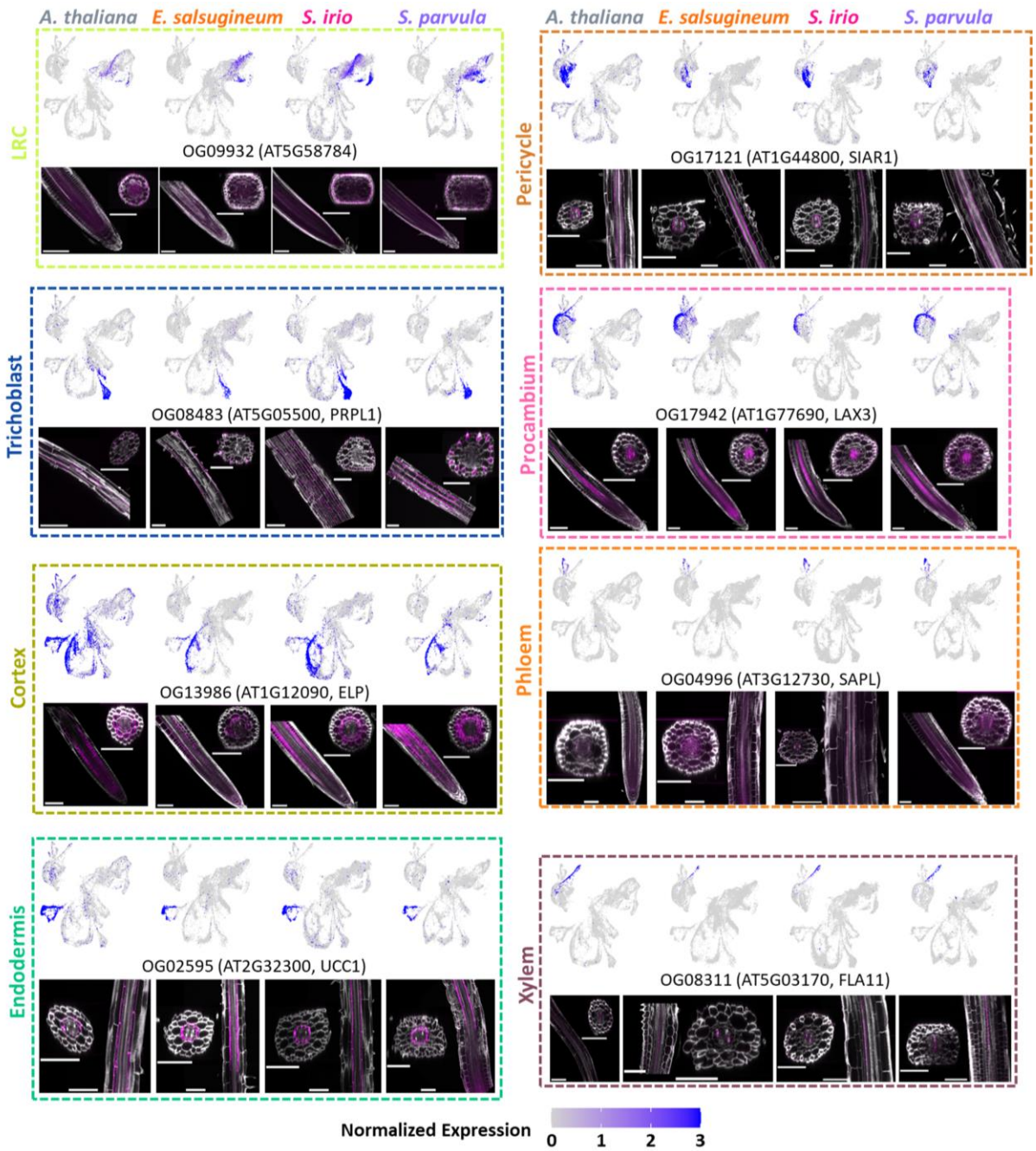


Figure S6 Whole-mount *in situ* HCR of cell type markers that are conserved between the 4 Brassicaceae species. Note that *S. irio* and *S. parvula* have two layers of cortex. PRPL1, proline-rich protein-like 1; ELP, extensin-like protein; UCC1, uclacyanin 1; SIAR1, siliques are red 1; LAX3, like aux1 3; SAPL, Sister of APL; FLA11, FASCICLIN-like arabinogalactan-protein 11. Scale bars, 100 μm.

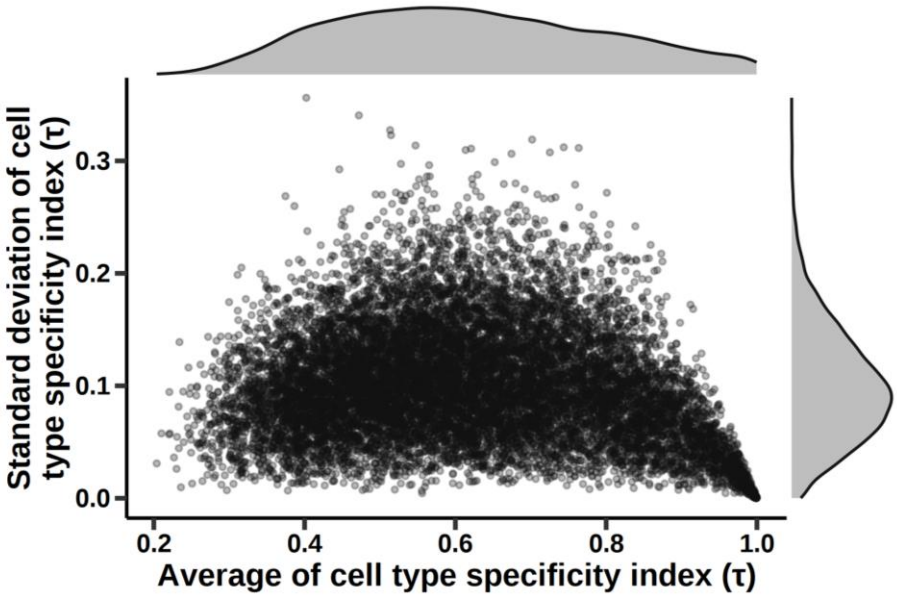


Figure S7 Comparison of the average and the standard deviation of cell type specificity index for the 1-to-1 orthologs. Cell type specificity index (τ) is based on tau metric of tissue specificity and calculated as $\sum_{i=1}^{N}(1-x_i)/(N-1)$, where N is the number of cell types and x_i is the expression 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.

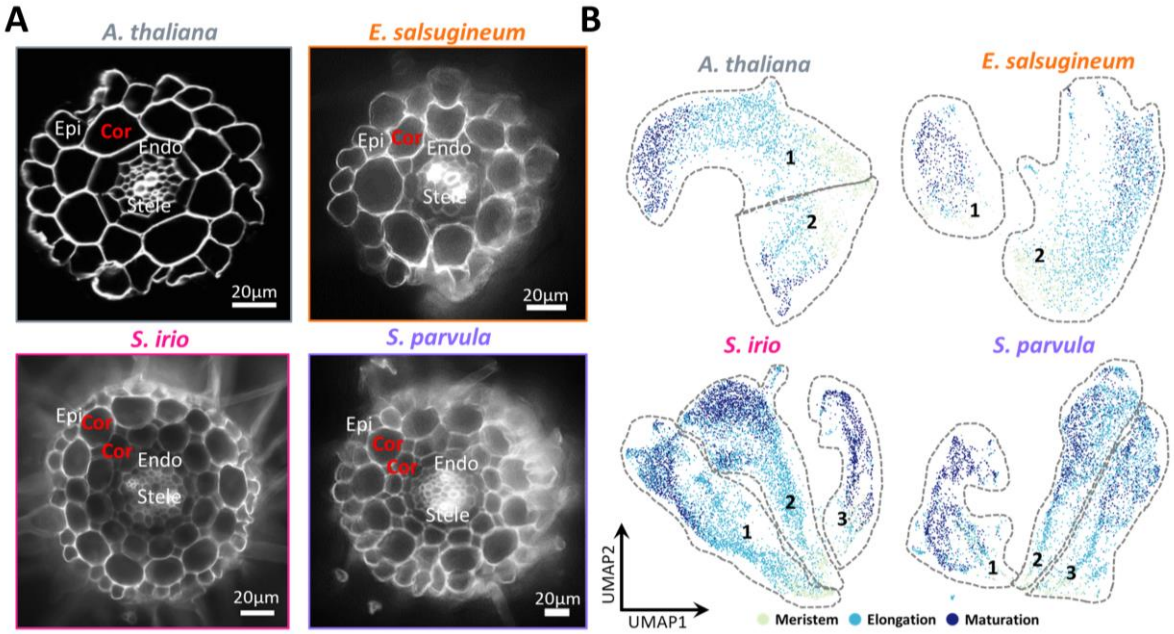


Figure S8 Comparison of cortex cells across species. (A) Cross-sections of 5–6-day old seedlings from each species. Scale bars, 20 μm. (B) UMAP representation of cortex cells in each species. Sub-populations in each species are circled.

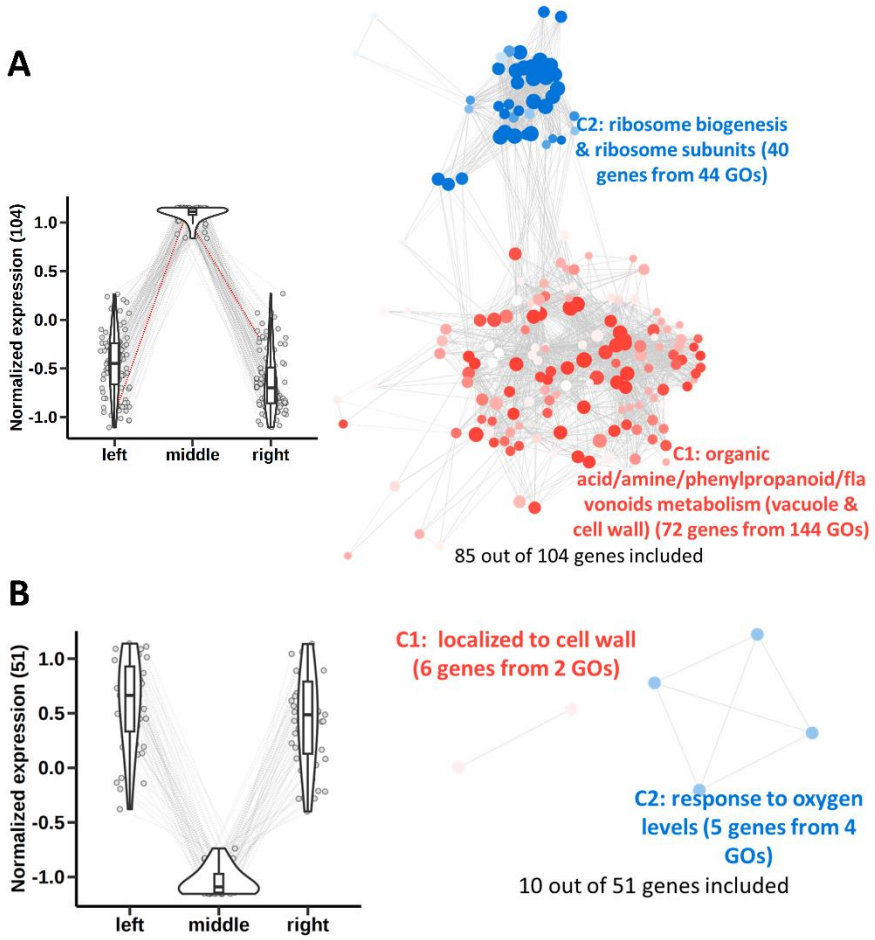


Figure S9 Functional representations of genes that were preferentially expressed (A) and depleted (B) in the middle cortex subpopulation of *S. irio*. The red dash line in the line graph in (A) marks the normalized expression of *Si-s1384-01650*, the HCR result of its expression in *S. irio* is shown in Figure 3D. 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 (false discovery rate (FDR)-adjusted $p < 0.05$) with darker shades indicating smaller P-values.

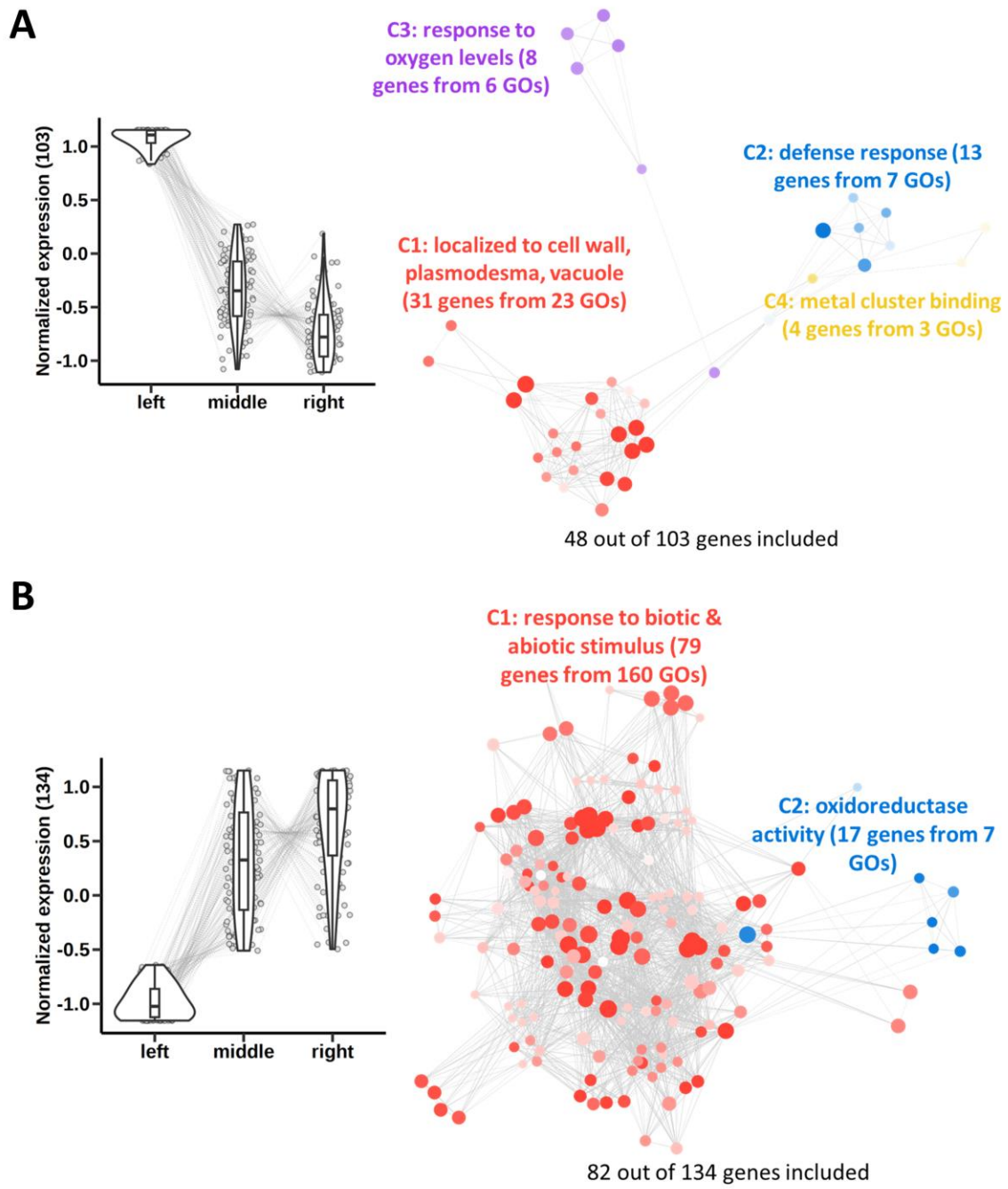


Figure S10 Functional representations of genes that were preferentially expressed (A) and depleted (B) in the left cortex subpopulation of *S. irio*. GO clusters are differently colored and labelled with the representative functional term. Each node represents a gene ontology (GO) term; node size represents 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 (false discovery rate (FDR)-adjusted $p < 0.05$) with darker shades indicating smaller P-values.

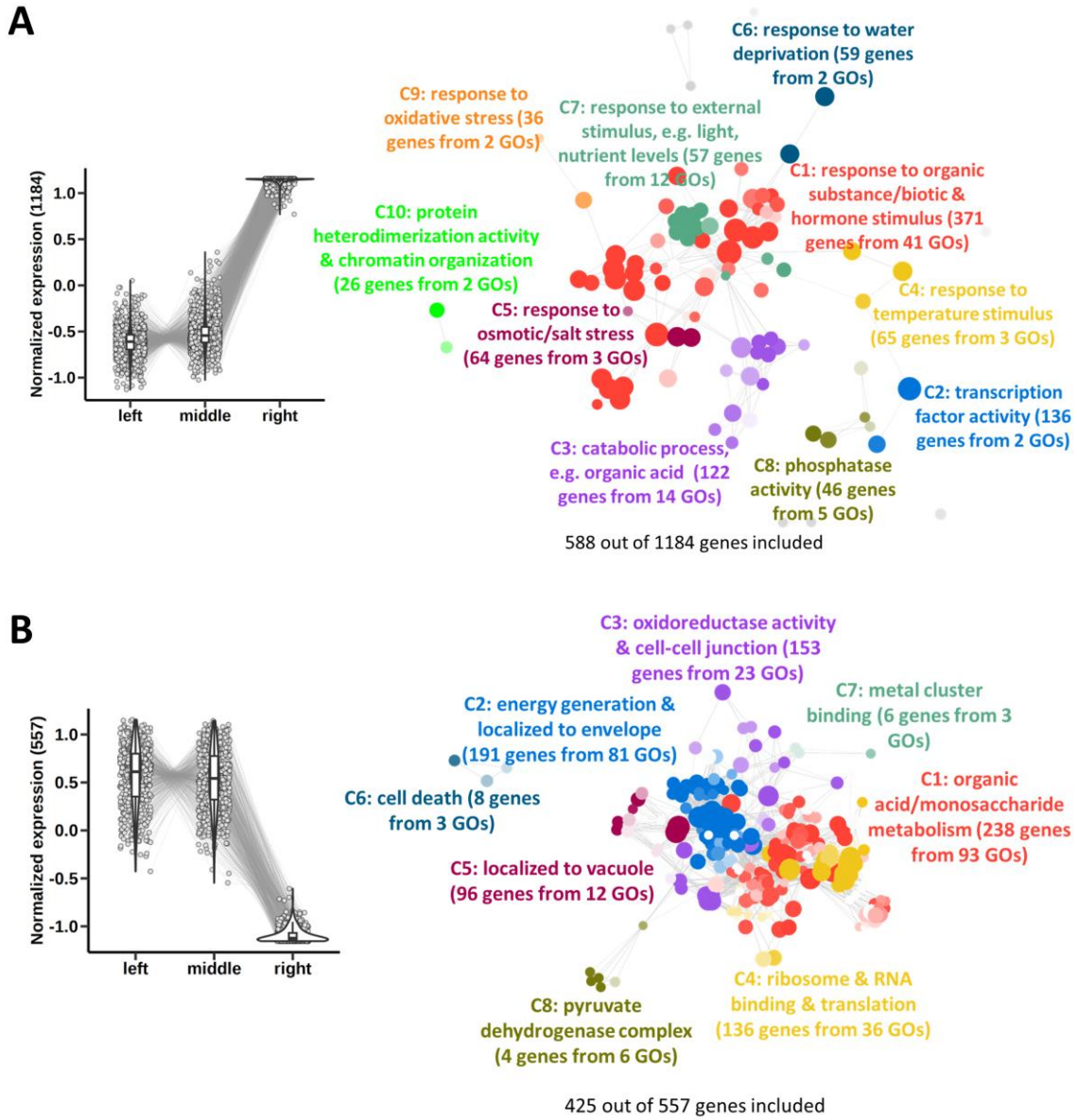


Figure S11 Functional representations of genes that were preferentially expressed (A) and depleted (B) in the right cortex subpopulation of *S. irio*. GO clusters are differently colored and labelled with the representative functional term. Each node represents a gene ontology (GO) term; node size represents 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 (false discovery rate (FDR)-adjusted $p < 0.05$) with darker shades indicating smaller P-values.

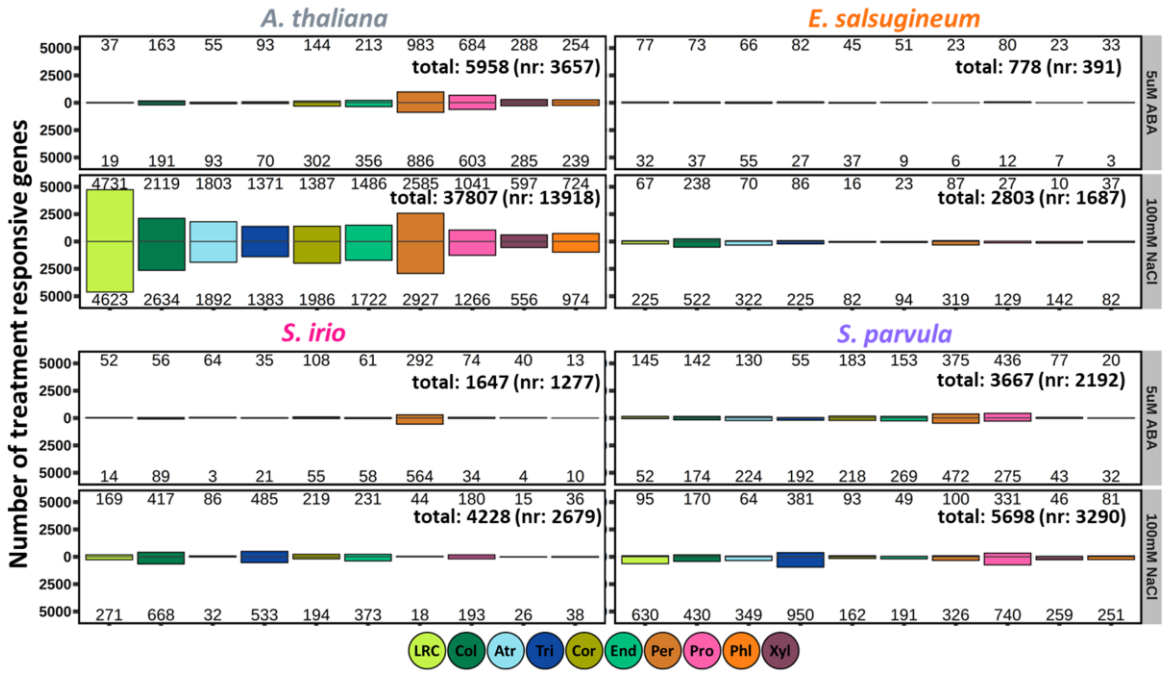


Figure S12 Number of stress responsive genes in each cell type from each species in response to 5 µM ABA (top) and 100 mM NaCl (bottom). Total number of stress responsive genes (nr, non-redundant) from all cell types in each species was shown on the top left corner.

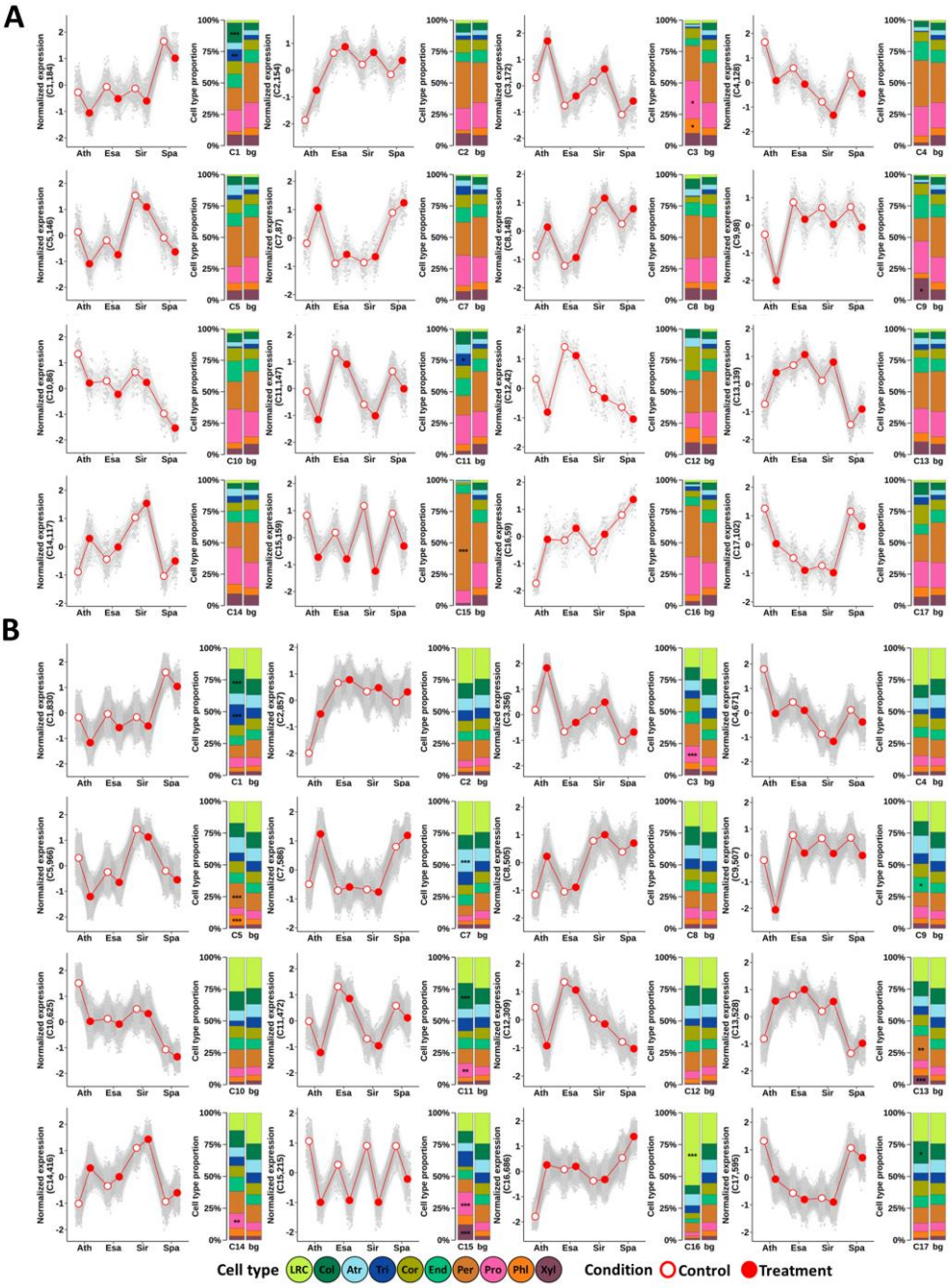


Figure S13 Contributions of each cell type to co-expression networks identified from stress-responsive orthologs under 5 µM ABA (A) and 100 mM NaCl (B). Cluster ID and the number of orthologs in the cluster are shown in parentheses in the Y-axis labels. Cell type proportions of the cluster were compared 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. Col, Columella; LRC, Lateral Root Cap; 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*.

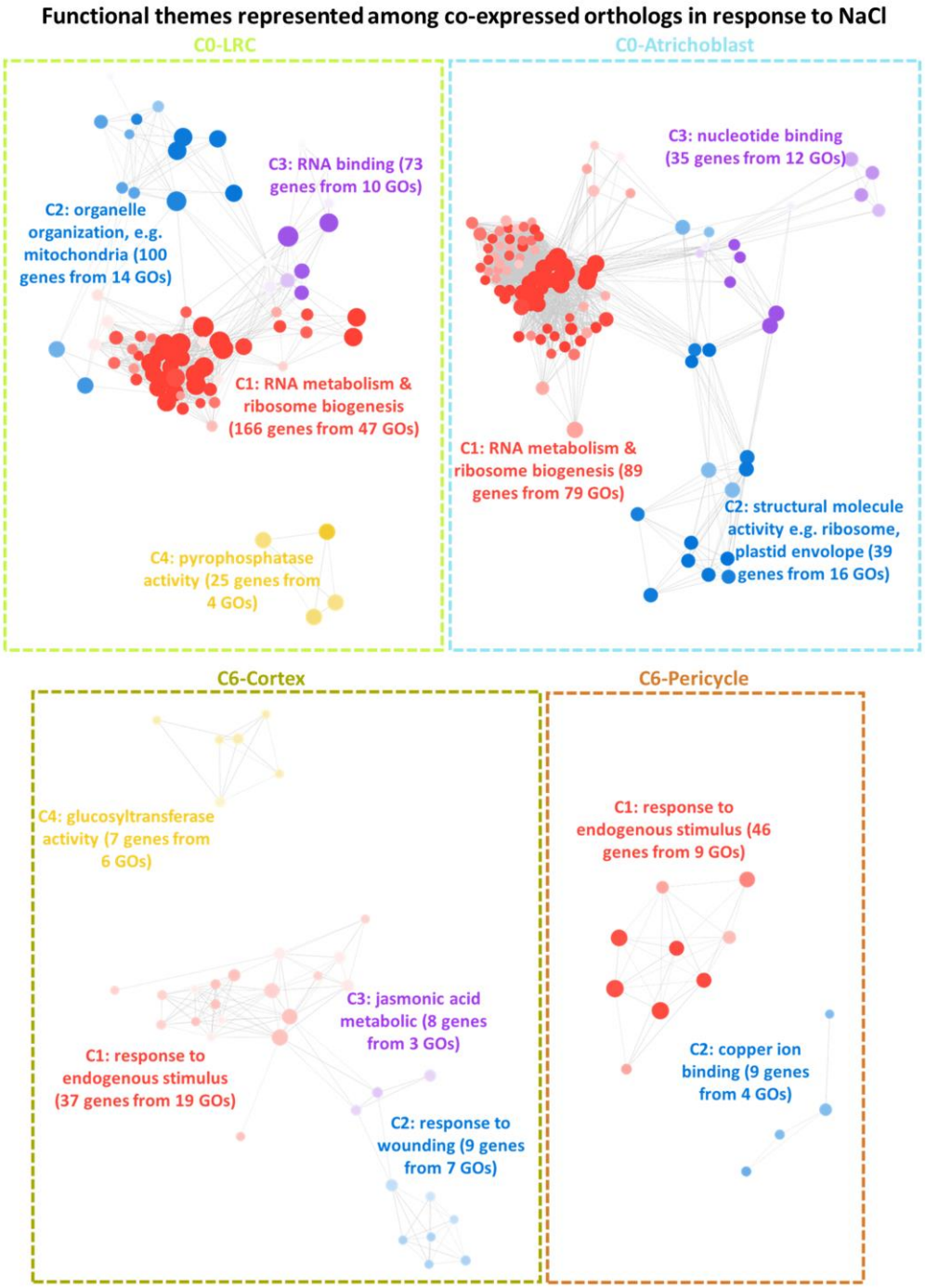


Figure S14 Functional representations of co-expressed orthologs in response to NaCl. GO clusters are differently colored and labelled with the representative functional term. Each node represents a gene ontology (GO) term; node size represents genes in the test set assigned to that functional term; GO terms sharing > 50% of genes are connected with edges; and the shade of each node represents the *p*-value assigned by the enrichment test (false discovery rate (FDR)-adjusted $p < 0.05$) with darker shades indicating smaller *p*-values.

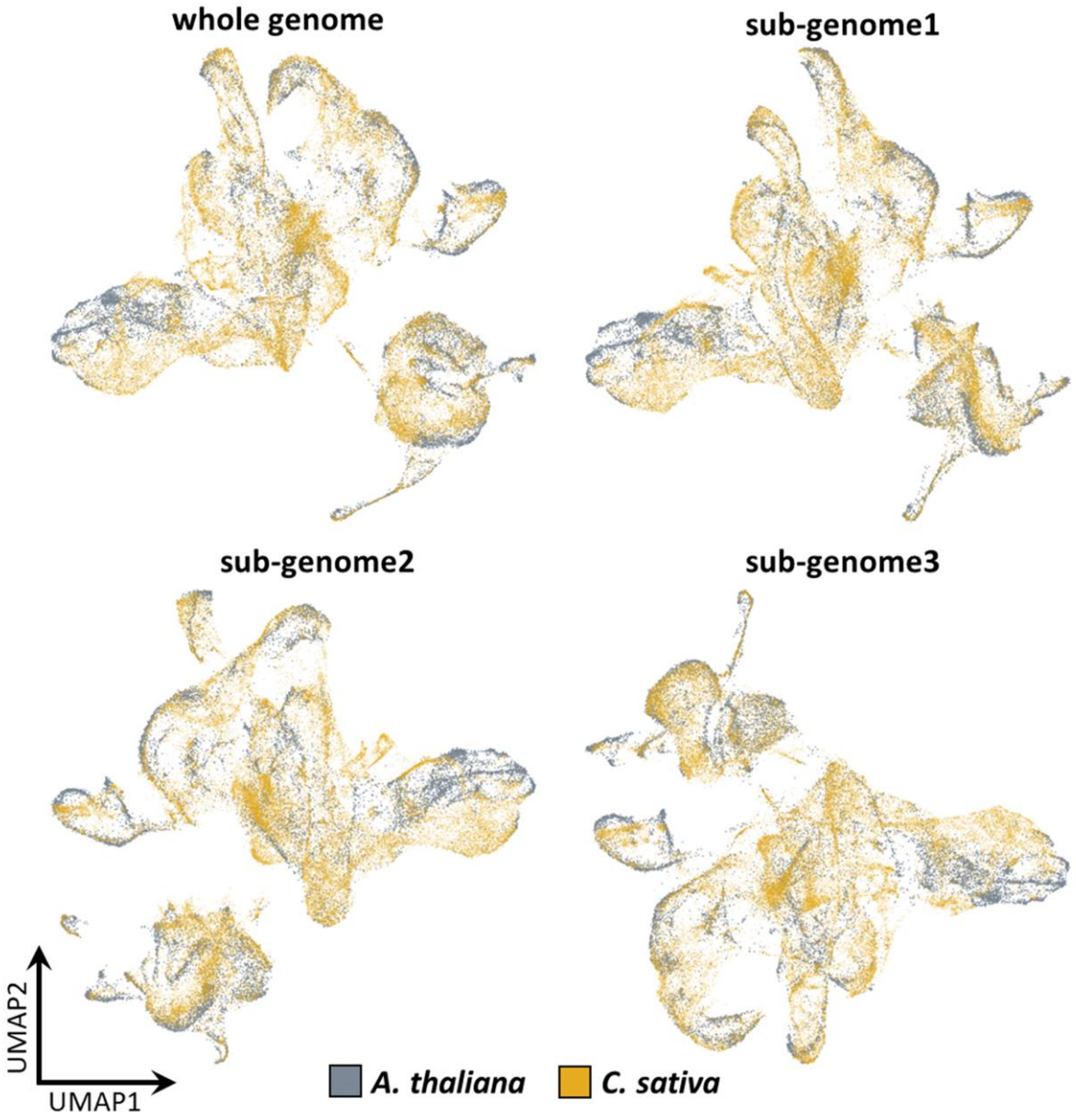


Figure S15 Integration between Arabidopsis and *C. sativa* using different sub-genomes in *C. sativa*.

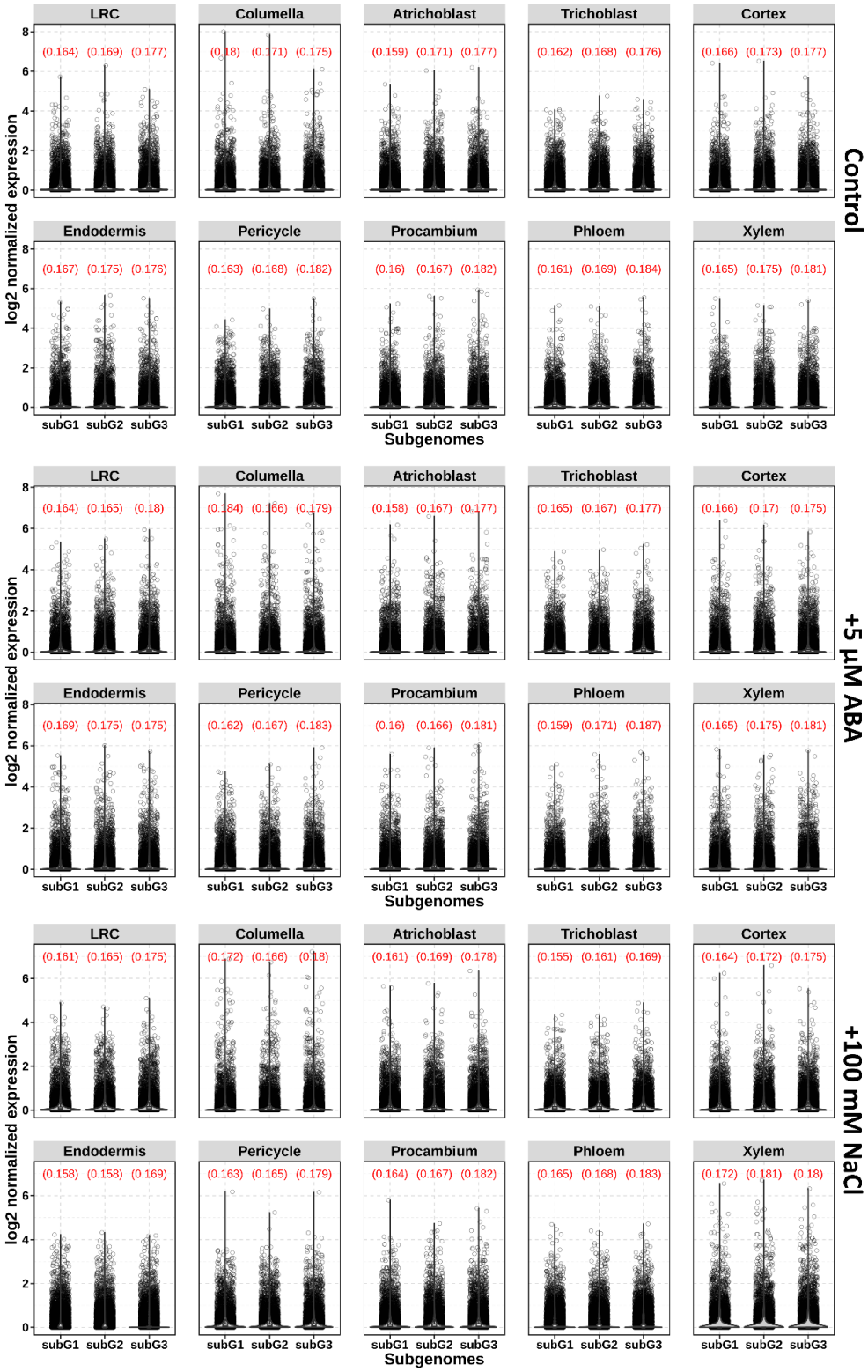


Figure S16 Overall transcriptome expression levels separated by the three sub-genomes for each cell type under different conditions. Numbers in red indicate the averages of transcript levels.

Figure S17 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 Table S6. The shade represents the *p*-value assigned by the enrichment test (false discovery rate (FDR)-adjusted *p*-value < 0.05) with darker shades indicating smaller *p*-values.

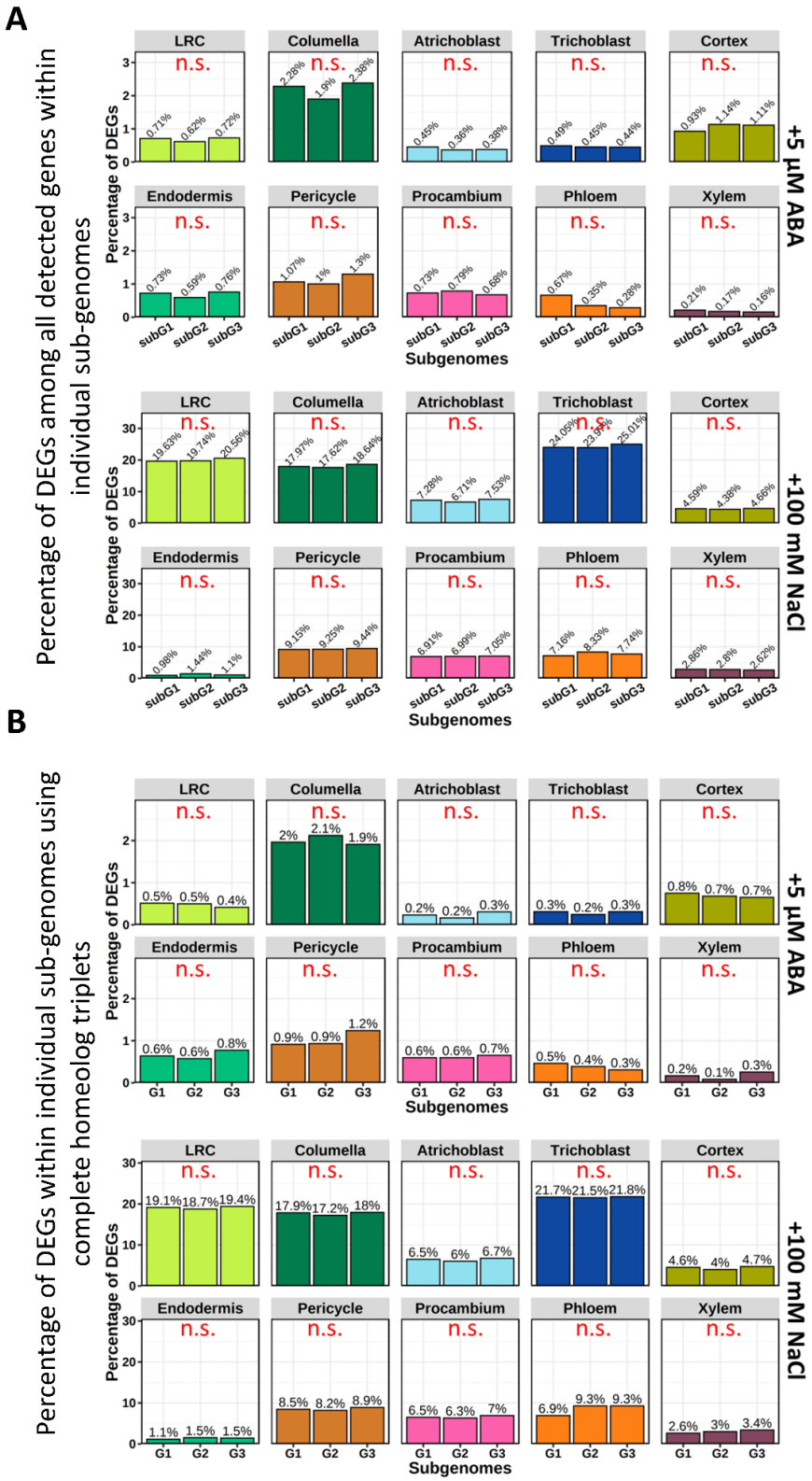


Figure S18 Percentage of DEGs in each sub-genome of *C. sativa* evenly distributed among the sub-genomes considering all detected genes within individual sub-genomes (A) and only the complete homeolog triplets (B). n. s. not significant based on hypergeometric test.

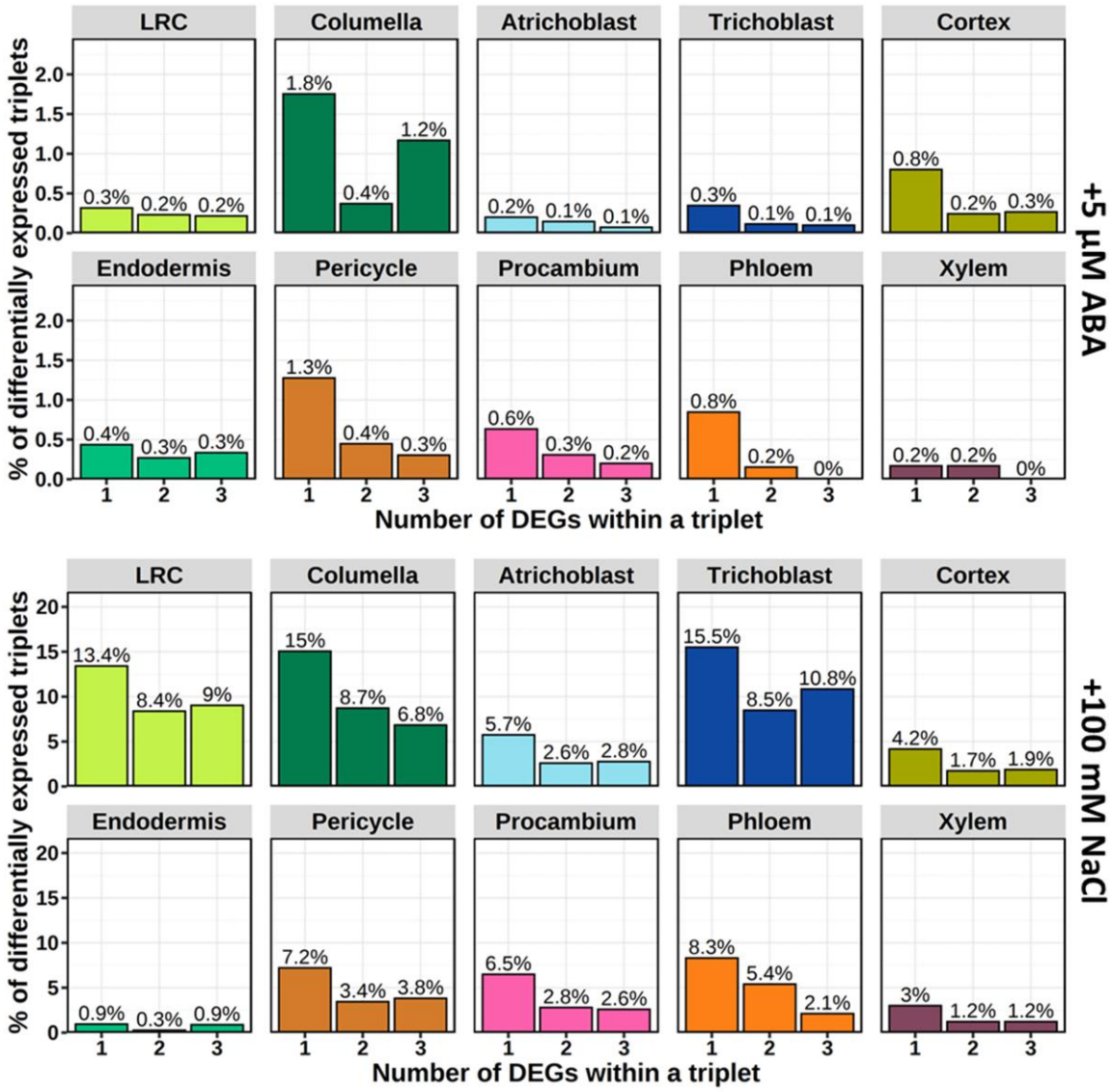


Figure S19 Percentage of triplets with 1, 2, or 3 stress responsive homeologs (DEGs).