

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

Corresponding Author: Dr Maheshi Dassanayake

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This study presents a comprehensive single-cell transcriptomic atlas of root cells from five Brassicaceae species—*Arabidopsis thaliana*, *Sisymbrium irio*, *Eutrema salsugineum*, *Schrenkiella parvula*, and *Camelina sativa*—under control, ABA, and salt stress conditions. Using 10X Genomics scRNA-seq, the authors generated over 174,000 root cell transcriptomes and developed a cross-species annotation framework (for 4 of the 5 species). They found that nearly half of *Arabidopsis* cell-type marker genes lacked conserved expression across species, prompting the identification of a new set of conserved Brassicaceae markers. The study revealed lineage-specific anatomical and transcriptional variation, especially in cortex subpopulations, and showed that stress responses are highly cell-type and species-specific. In *Camelina* sub-genomes seemed to contribute equally to stress responses, and divergence among homeologs was largely uncoupled from coding sequence variation. While the study identifies divergent expression patterns, functional validation of these genes is limited, however the validations that were done seem reasonable. I am not expecting the authors to do more validations. A few details are missing that I hope could be included or expanded upon, however the work is very solid, and the authors are using state of the art methods for analysis- especially the prediction algorithms, I have no major notes on types of analysis. Because of the nature of stress (reduction in many types of cell-type expression, see Jean-Baptiste heat shock single-cell), a major concern would be that more abundantly sampled cell types would have the strongest signal simply due to power. However, the authors do their best in compensating for that in several ways. The *Camelina* work could use more expansion and feels very much like an afterthought or even a completely different paper, loosely related to the expansive, integrated analysis in the rest of the paper. After addressing the relatively straight forward requests below I can only suggest publication so strongly.

- Quiescent center cells were underrepresented (really, completely absent), limiting insights into stem cell niche dynamics. I would love to hear more about why this is, most other studies in *Arabidopsis* (even those not using root tips) have at least some QC present. The only mention is just that it is under-sampled, but with 100,000+ cells I'm not sure this is the only reason.
- The analysis of sub-genome dominance in *Camelina* is largely descriptive and lacks mechanistic depth. There is very little description of what has to be a challenging problem, dealing with hexaploidy genes, was the alignment rate drastically less? Were there particular alignment strategies? Etc. The methods do not have any of these types of details either. Presumably these types of challenges are why *Camelina* was singled-out and not included or attempted to be included in the cross-species analysis, but this is not addressed as far as I read.
- The overall UMIs are on the lower side, could you please provide knee plots for these experiments in the supplement.
- There is a period missing on line 2 of the abstract (line33).
- Perhaps you could further expand on line 204-205, when you say "This discrepancy may be due to the technical differences between the two complimentary RNA detection methods"
- I also wonder if some of the cell type differences are simply like structural/lifestyle differences, where the actual physical differences of the roots leads to different penetration etc. of the media stresses.

Josh Cuperus

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

In this interesting manuscript, Wang et al., explored the diversity of root cell types and their expression patterns across the Brassicaceae family under salt and ABA stresses. The authors performed a multi-species single-cell comparative transcriptomics of the root, first to identify conserved and non-conserved cell type gene markers. They found that almost half of Arabidopsis gene markers are not conserved; however, through evolutionary and comparative analyses, they defined a set of conserved markers. Second, they identified transcriptomic evidence that supports the specialization of cortical cell subpopulations in *S. parvula* and *S. sirio*. Third, they observed a divergence in stress responses among species, especially in response to salt stress. Fourth, one of the most interesting contributions in this manuscript is the stress response between orthologs, trying to link the transcriptomic shifts to phylogenetic relationships, confirming the cell type contributions to the stress-prepared transcriptomes in extremophytes. Finally, they used the *Camelina Sativa* single-cell RNA-seq data to explore the effects of polyploidy on cell-type specificity.

Overall, this manuscript is of great significance to the root biology community, serving as a perfect complement to the root transcriptomes of numerous plants at single-cell resolution. This enhancement will undoubtedly provide valuable resources to the plant community. While this manuscript has the potential to be published in the journal, it requires further clarification and restructuring. I have attached my comments below for your reference.

Major:

1) One of the most important results of this manuscript is that the authors identified half of the markers for Arabidopsis as not conserved across the Brassicaceae family. My concern here is that this result is a reflection of the low gene expression because of the library size or the plant species. How can we be confident that these "lost" of expression is not noise across the gene expression between species? Could the authors clarify in the main text and methods the normalization of the expression per library size, plant species, and batch effects? Moreover, I was wondering what happened with the other half of the Arabidopsis markers. Are they entirely conserved across the Brassicaceae family?

2) I found interesting the cortex specialization in *S. parvula* and *S. Sirio*. Which are the top-expressed genes in each cortex layer? What is the transcriptomic signature or co-expressed genes for that extra cortical layer? More importantly, what happens if you extend this comparative analysis of cortical cells using available data for tomato and *Medicago truncatula* (two species with two layers of cortical cells)? Are the same genes involved? I suggest adding at least these cortical cell data of the root to extend the results and generalize about this specialization process in roots.

3) I'm concerned about the restructuring of the paper. In line 93, the authors described the number of cells for each plant species (5 root species), but surprisingly, in Fig. 1, only 4 plant species appear. What happened with *Camelina sativa*? Are there some contradictions. The authors also mentioned that they generated between 36,000 and 50,000 root cell transcriptomes for the five plant species. However, as shown in Fig. Supp 1, *Camelina sativa* has 27,442 cells. I can imagine that could be related to the restructuring of the manuscript, because at the end of the manuscript, the last section (involving Fig. 7) is difficult to follow, with likely a disconnection of the main story. Starting again with a UMAP, and those results are described at the beginning.

4) Figure 5 is challenging to comprehend. Panel A is too small. Panel B illustrates the transcriptome response (I can envision differential gene expression) for each cell type and treatment. Here, it's crucial to quantify the number or percentage of cells responding to each stress in each plant species. In this way, the authors can identify the cell types with the most specific response. Please clarify panel E.

5) The title suggests that the study focuses on the evolutionary diversity and stress response in roots. However, this study is in the Brassicaceae family.

Minor:

6) Line 33. typo: "yet cross-species comparisons at single-cell resolution remain scarce. To explore this diversity"

7) Clarify the statistical process for normalizing pseudo-bulk samples across libraries and species.

8) The results section for the abstract and introduction needs more detail and clarification; the descriptions are very general.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

In this manuscript, Wang, Ryu, Dinneny, and colleagues used scRNA-seq to create a cross-species transcriptomic atlas of Brassicaceae roots treated with abscisic acid and NaCl. The authors found that the expression of a majority of Arabidopsis thaliana root cell type markers are not conserved in the other profiled species. Thus, validation of newly identified markers via HCR produced a toolkit of conserved markers and revealed orthologs with lineage-divergent expression profiles. Taking advantage of their atlas resource, the authors investigated how different cell types respond to stress between species and across treatment types.

Overall, I believe that the computational framework established for cross-species and pan-treatment analysis of scRNA-seq data will surely be a useful resource for the plant biology community. However, to me, the significance of the biological insights in the manuscript are less clear. I will expand on these thoughts and provide additional comments below.

1. Although the authors undertook a comprehensive annotation of root cell types, there is no mention of the different developmental zones of the root, save for the labels in Figure 3, panels E and F. I assume that the data captures primary root

tips through at least the early differentiation zone, but there is no indication of how much of the root was harvested for scRNA-seq for each species (lines 551-552). Further, although different parts of the root (e.g., meristem versus elongation and/or differentiation zones) are shown in Figure 2 to correspond to expression of the different marker genes, it is unclear if these differences in spatial localization match the scRNA-seq data.

Can the authors please clarify how much of the root was harvested for each species, what developmental zones were captured in the scRNA-seq data, and how they determined in which developmental zone the putative marker genes were expected to be expressed in the HCR experiments? Additionally, can the authors address if markers expressed in specific developmental zone(s) will be useful for understanding stress responses (e.g., is it an open question if cells more responsive to stress depending on their differentiation state)?

2. Figure 4; Lines 211-253; Supp Fig 13. To me, this section seems out of place with the stress-response focus of the rest of the manuscript. Can the authors provide context for the significance of a cortex layer that specializes in flavonoid metabolism? Are flavonoids important for stress responses or plant resilience?

3. Figure 5, lines 325-338. The identification of cell type transcriptional profiles that diverge across species and treatments is interesting but the biological context of these results is somewhat lacking. For example, the authors note that procambium cells showed the most increases in transcriptomic divergence between species, particularly in response to ABA (lines 336-337). Is the identification of procambium particularly interesting compared to other cell types? Is procambium function known to differ across species?

4. Figure 6, lines 392 – 419. The authors investigated whether the previously reported stress-prepared state at the organ level in extremophytes is due to specific cell types. In response to salt treatment, the authors found that outer layer cells (LRC and atrichoblasts) contribute to high constitutive gene expression in the profiled extremophyte species. They conclude that this result suggests a tendency for some cell types to play a more substantial role than others in shaping the stress-prepared signal. Could the authors address if this response could be due to the fact that the outer layer of cells is directly exposed to the environment and to salt treatment in the experiment? Couldn't it therefore be expected that they would respond differently than other cell types?

5. Figure 7. The authors demonstrate that their cross-species cell type annotation strategy enables successful annotation of scRNA-seq data from *C. sativa*. This analysis revealed a lack of large-scale sub-genome dominance in stress responses (Lines 450-453). I don't fully understand the implications of this conclusion. Could the authors clarify how sub-genome dominance (or lack thereof) in stress responses is useful for or a hindrance to plant resilience?

6. I would like to better understand how the authors envision that the community will use the new resources developed for these Brassicaceae species. For example, beyond confirming the cross-species annotation, will the newly validated cell type markers enable new downstream experiments?

Minor

1. On the Shiny app, 'atlas' is misspelled in the title at the top of the page.

2. Supplemental Fig 10: It seems like part of the final sentence in the figure legend was omitted.

3. The processed data deposited to GEO are currently under embargo. Could the authors confirm if annotated Seurat objects are available for the atlas? If not, I suggest providing them to save time and computational resources for the community.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)

his manuscript presents a multi-species single-cell RNA-seq atlas of Brassicaceae roots under control, ABA, and salt stress conditions. The authors profile ~175k cells from *Arabidopsis*, *Sisymbrium irio*, *Schrenkiella parvula*, *Eutrema salsugineum*, and *Camelina sativa*, identify 10 major root cell types across species, and build a comparative framework based on ~15k 1-to-1 orthologs. They show that nearly half of classical *Arabidopsis* root cell-type markers are not conserved in expression specificity in the other species and define a new set of conserved Brassicaceae markers validated by HCR in situ hybridization.

Using phylogenetically informed analyses, they quantify divergence of cell-type-specific expression among orthologs, highlight lineage-specific shifts and extremophyte vs non-extremophyte signatures (e.g. NIP5;1 and GSTU25), and analyze how ABA and NaCl responses redistribute across cell types and species. They further examine the allohexaploid *C. sativa*, showing partial expression dominance of one sub-genome but largely balanced contributions to stress responses and uncoupling of coding versus expression divergence among homeolog triplets. Finally, they provide an interactive browser and processed data matrices as a community resource.

Overall, this is a substantial and timely resource for comparative plant single-cell biology with broadly interesting biological insights into cell-type evolution and stress adaptation.

Major comments:

Figure 5A: Why did the authors not use the UMAP analysis, but PCA? Please justify the choice of the methods. How come the different cell type from the same species show a smaller distance than the same cell types from the different species? I would expect at least some cell types to show a highly conserved expression, and consequently, a shorter distance within the PCA. This makes me question the validity of the method, and whether the observation that the two extremophytes are functionally related. This is also suggested in Figure 5C, where the *S. parvula* and *S. irio* are now more related (due to phylogeny).

Minor comments:

L110: why is ABA concentration given as 'micromolar'

Figure 5D: I assume that the columns for each species correspond to the cell types in the previous panels. Please make this clear

L324: 'We observed 324 overall larger transcriptomic distinctiveness between species in response to 100 mM NaCl 325 compared to that of 5 μ M ABA (Fig. 5E).' I am unsure how the authors came to this conclusion from the PCA figure. Please make this more clear.

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)

In this study, Wang et al. generated comprehensive and valuable single-cell RNA-seq datasets from Brassicaceae roots, including stress-sensitive species, extremophytes, and a polyploid crop, under NaCl and ABA treatments. They also developed a framework for comparative analyses using scRNA-seq datasets. While the study provides insights into how stress responses diversify across cell types and species, several concerns remain:

1. Estimation of divergent responses: The authors rely heavily on previously published methods to infer gene expression divergence among species. However, it is unclear exactly how this was done and why specific thresholds (e.g., probability cutoffs) were chosen for further analyses. I also wonder whether the estimation approach, originally developed for mammalian cells, can be directly applied to crops.

2. Lack of quantitative measurements: Although the authors provide the numbers of orthogroups in most figures (e.g., Figures 1E and 3B), they proceed directly to discussing individual genes. This makes it difficult to assess the extent or proportion of genes affected or belonging to each group. For instance, in Figure 5F, the authors state that "cortex and atrichoblast transcriptomes significantly increased their similarity following NaCl treatment only between the two extremophytes." However, without more quantitative data, this is not clearly supported by the current visualizations.

3. Scope of ortholog comparisons: In most analyses, the authors focus on 1:1 orthologs (15,198), except in Figure 6, where at least 30,133 genes were included. This suggests that many claims are based on only one-third to one-half of the total genes in each species. It would be important to assess whether genes with one-to-many or many-to-many orthologous relationships show similar trends.

4. Interpretation of species and treatment effects: In the abstract, the authors claim that "cell-type contributions to stress responses differed both between species and across treatments." However, based on Figures 5 and 6, this conclusion appears largely descriptive and not strongly supported by comparative analyses. Specifically, the NaCl and ABA datasets were analyzed separately rather than jointly. Given the well-known cross-talk and overlapping responses between salt and ABA signaling, it would be valuable to identify genes showing shared or distinct responses to these two stresses across species and cell types. For example, a co-expression clustering approach combining both stress conditions could strengthen this point.

5. Analysis of ABF and ABA responses: As noted by the authors, ABF directly binds to ABA-responsive elements, so the overrepresentation of ABA-related events compared with salt treatment is not surprising. The authors argue that this may reflect a more complex ABA-mediated regulation during salt stress (Figure 6C), but this remains speculative. Performing DAP-seq with salt-specific transcription factors could provide a meaningful counterpart analysis. Additionally, the statistical test described in the Methods (lines 810–811: "The significance of differences in the percentage of divergently regulated orthologs between the two defined groups were determined with t.test function in R") may not be appropriate, as t-tests assume normal distribution.

6. Shiny app performance: Although the authors acknowledged long loading times, the current Shiny apps are still difficult to use. Despite multiple attempts with different internet connections and browsers, I repeatedly encountered timeout errors.

(Remarks on code availability)

It's pretty clear to me.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Authors have addressed my modest concerns. Please publish.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

Reviewer #2

Wang et al., explored the diversity of root cell types and their expression patterns across the Brassicaceae family under salt and ABA stresses. Overall, this manuscript holds great significance for the root biology community and will be a very good resource. It serves as an excellent complement to the root transcriptomes of numerous plants at single-cell resolution. After reviewing the revised version, the authors addressed all my concerns and questions. Therefore, this manuscript will be a valuable addition to this journal.

Below are the first round questions and answers, and my answers to the revised version.

Major:

First round 1) One of the most important results of this manuscript is that the authors identified half of the markers for *Arabidopsis* as not conserved across the Brassicaceae family. My concern here is that this result is a reflection of the low gene expression because of the library size or the plant species. How can we be confident that these “lost” of expression is not noise across the gene expression between species? Could the authors clarify in the main text and methods the normalization of the expression per library size, plant species, and batch effects? Moreover, I was wondering what happened with the other half of the *Arabidopsis* markers. Are they entirely conserved across the Brassicaceae family?

First round author response: We thank the reviewer for this important concern and agree that it is essential to distinguish true evolutionary divergence from technical noise.

In our analysis, cell-type markers were identified using Seurat’s FindMarkers within each species, where markers were defined as genes showing consistent and cell-type-enriched expression relative to their expression in other cell types in the same species. Thus, by definition, markers represent genes with sufficient expression levels to be reliably detected and enriched within a given dataset. To examine the expression of these cell-type markers across species, we used a cell-type pseudo-bulk strategy, in which raw gene counts in each cell were normalized by the total counts of that cell and the resulting normalized values were averaged across all cells of the same cell type. This approach captures the average per-cell transcriptional state while minimizing effects of cell number and library size differences. We have clarified these details in the Methods at lines 818-823.

When assessing conservation across species, we observed that genes classified as “nonconserved” were not universally lowly expressed. Instead, these genes often showed substantially detectable expression and clear marker behavior in one or more species, while being lowly expressed in other species, where they therefore could not be identified as valid markers (Supplementary Fig. 9). Importantly, orthologs of these genes in other Brassicaceae species frequently retained robust expression and cell-type specificity. Thus, “non-conserved” refers to species-specific attenuation or loss of marker-level expression, rather than technical dropout or global loss of expression.

With respect to the remaining half of *Arabidopsis* markers, these genes retained consistent expression levels and cell-type specificity across all Brassicaceae species examined and were therefore classified as conserved markers (Supplementary Fig. 9). This list represents the most robustly conserved root cell-type markers currently known for Brassicaceae species, although this list will likely be refined and expanded as additional root single-cell transcriptome datasets become available for other members of the family. We have clarified these details in the Results at lines 155-160.

Second round reviewer: Thank you to the authors for the methods clarification. I am convinced now that the term “non-conserved” could be the loss of marker-level expression more than noise.

First round reviewer 2) I found interesting the cortex specialization in *S. parvula* and *S. Sirio*. Which are the topexpressed genes in each cortex layer? What is the transcriptomic signature or co-expressed genes for that extra cortical layer? More importantly, what happens if you extend this comparative analysis of cortical cells using available data for tomato and *Medicago truncatula* (two species with two layers of cortical cells)? Are the same genes involved? I suggest adding at least these cortical cell data of the root to extend the results and generalize about this specialization process in roots.

First round author response: We thank the reviewer for raising these important questions. We have performed the following analyses to address these questions.

To identify the transcriptomic signatures of the inner and outer cortex layers, we compared the transcriptomes of different cell layers (with cortex divided into inner and outer layers) in *S. irio*, and identified the genes co-expressed for the different cortex layers (Supplementary Fig. 17).

We found that genes involved in water transport and environmental sensing were highly expressed in both inner and outer cortex layers compared to other cell types (Supplementary Fig. 17). Genes preferentially expressed in the inner cortex were mostly associated with secondary metabolism (particularly flavonoid biosynthesis), cell-to-cell communication via

plasmodesmata, and jasmonic acid and ABA signaling. In contrast, the outer cortex prioritized the expression of genes associated with hypoxia sensing, thermal stress response and gibberellin signaling pathways (Supplementary Fig. 17).

To investigate the cortex specialization in other species, we extended our analysis to other species with more than one cortex layer in roots and corresponding root single cell transcriptome data, including rice⁶ and Medicago⁷. We did not include tomato as we were not able to locate a sufficiently-annotated dataset for this species. We extracted cortex cells from rice and Medicago and applied the same within-species clustering approach to these cells to maximize the sensitivity for detecting internal subpopulation structures. We identified two major subpopulations in the rice cortex cells, corresponding to the elongation and maturation zones based on the developmental zone annotation (Supplementary Fig. 18). Notably, the rice cortex cells exhibited a single, linear progression from elongation to maturation, in clear contrast to the branching paths we observed for cortex cells in *S. irio* and *S. parvula* in our study. This lack of distinct branching trajectories suggests that, at this resolution, rice cortex cells maintain a relatively unified transcriptional identity as they transition toward their terminal functional states, with developmental stage being the primary driver of transcriptomic variance rather than radial cell-layer specialization. We also detected the subpopulation structure for the Medicago cortex cells (Supplementary Fig. 18), however, we were unable to clearly determine the biological relationship of the subpopulations due to the lack of reliable developmental annotation for the Medicago dataset, preventing a definitive comparison of its trajectory direction. It is important to note that because we primarily examined cells from stress-neutral conditions, we cannot rule out the possibility that cortex specification in these species is condition-dependent or inducible upon environmental changes.

To further investigate whether this transcriptomic uniformity reflects a lack of functional specialization, we examined the expression of flavonoid biosynthesis genes, which are differentially partitioned between different cortex layers in our data. Our analysis demonstrated that these genes are either undetected, including CHS which catalyzes the first committed step in the flavonoid biosynthetic pathway, or universally expressed across all cortex cells, rather than being limited to a specific subset of cortex cells (Supplementary Fig. 18). This absence of partitioned metabolic signatures reinforces our subpopulation analysis results and suggests that, even if cortex specification does exist in these species, the flavonoid pathway is not involved in that specialization. We have added the results of these additional analyses to the revised manuscript at lines 256-266 and 278-285 (the corresponding method is added at lines 957-963).
Second round reviewer: Many thanks to the authors for performing many analyses regarding the cortex specialization and adding rice and Medicago.

First round 3) I'm concerned about the restructuring of the paper. In line 93, the authors described the number of cells for each plant species (5 root species), but surprisingly, in Fig. 1, only 4 plant species appear. What happened with *Camelina sativa*? Are there some contradictions. The authors also mentioned that they generated between 36,000 and 50,000 root cell transcriptomes for the five plant species. However, as shown in Fig. Supp 1, *Camelina sativa* has 27,442 cells. I can imagine that could be related to the restructuring of the manuscript, because at the end of the manuscript, the last section (involving Fig. 7) is difficult to follow, with likely a disconnection of the main story. Starting again with a UMAP, and those results are described at the beginning.
First round author response: We thank the reviewer for raising this important point regarding the clarity and organization. We agree that the current presentation could be difficult to delineate and that the handling of the *C. sativa* dataset requires additional clarifications.

As described earlier in response to Reviewer 1, because of the challenge in integrating polyploid species with diploid species, we intentionally divided our manuscript into two related but distinct sections: 1) a comparative single-cell analysis of four diploid Brassicaceae species, and 2) a separate application of the same framework to the polyploid crop *C. sativa*. This is because all our cross-species comparisons rely on 1-to-1 orthologs, and including the hexaploid *C. sativa* introduces a complex 1-to-3 orthology relationship. Therefore, we focused Fig. 1-6 exclusively on the comparisons among the four diploid species. *C. sativa* is introduced later as a dedicated case study to evaluate cell-type annotation transfer and sub-genome contributions in a polyploid context as shown in Fig. 7. We have revised the manuscript to explicitly explain the layout of our manuscript and frame *C. sativa* as a separate extension of the analytical framework, rather than a continuation of the main diploid comparison.

Regarding the cell numbers, the statement that "36,000 to 50,000 root cell transcriptomes were generated for each species" refers specifically to the four diploid species. We've separately profiled over 27,000 cells for *C. sativa*, which is lower due to differences in protoplast recovery and sequencing yield. We have clarified this in the revised manuscript at lines 105-107 and 114-117.

Second round reviewer: After the explanation, the manuscript structure makes more sense. Thank you for revising the manuscript and providing a better explanation of the layout.

First round 4) Figure 5 is challenging to comprehend. Panel A is too small. Panel B illustrates the transcriptome response (I can envision differential gene expression) for each cell type and treatment. Here, it's crucial to quantify the number or percentage of cells responding to each

stress in each plant species. In this way, the authors can identify the cell types with the most specific response. Please clarify panel E.

Response: We thank the reviewer for this helpful feedback and agree that Fig. 5 requires clearer presentation and detailed explanations. We have enlarged Fig. 5a for better readability.

Fig. 5b does not represent the number or percentage of cells responding to the treatments, but rather the degree of transcriptomic separability between control and treated cells within each cell type in each species, as quantified by the Augur framework⁸. Augur measures the degree to which treated cells can be distinguished from untreated cells based on their overall transcriptomic profiles, yielding an AUC score for each cell type. Cell types in which treated and control cells are easily separable (reflected as having a high AUC score) are interpreted as having stronger transcriptional responses or shifts, whereas cell types in which treated and control cells remain similar (reflected as having a low AUC score) are interpreted as having weaker responses or being unchanged upon exposure to stress. The results reflect the magnitude of transcriptomic change for a cell type in response to the treatment, rather than the number or proportion of responding cells. Because of the high sparsity in scRNA-seq data and also because we've used "pseudo-bulk" differential expression analysis to identify differentially expressed genes as it has been shown to achieve the highest fidelity^{9,10}, it is currently impossible and not reliable to directly quantify the number or the percentage of cells responding to stress. However, we have provided the number of differentially expressed genes per cell type and treatment in each species in Supplementary Fig. 23 and Supplementary Data 6, which offers a complementary and more intuitive measure of cell type-specific responsiveness.

Fig. 5e aims to determine which cell type shows the largest transcriptomic divergence across species and the impact of treatments on the cross-species transcriptomic divergence. Each data point represents the transcriptomic divergence between six pairwise species comparisons (shown as the x-axis in Fig. 5f) under a given condition. On the PCA space shown in Fig. 5e, columella cells from different conditions form a unique cluster that is more separated from the rest of the cell types on PC1. This is due to columella cells exhibiting the largest transcriptomic divergence across species (Fig. 5f and Supplementary Fig. 27), suggesting that the columella cells from these species are highly dissimilar to each other, aligned to what has been reported for grass family species¹¹. We have also added a supplementary figure (Supplementary Fig. 26) to quantitatively measure the distances between treated and control samples.

Second round reviewer: Thanks for the detailed explanation of figure 5.

First round 5) The title suggests that the study focuses on the evolutionary diversity and stress response in roots. However, this study is in the Brassicaceae family.

First round author response: We thank the reviewer for raising this point. We agree that the study is focused on species within the Brassicaceae family, and our intent with the title was not to imply evolutionary comparison across all plant lineages. To avoid any possible ambiguity, we have revised the title to explicitly reference the Brassicaceae family at line 1.

Second Round reviewer: Thanks.

Minor:

First round 6) Line 33. typo: "yet cross-species comparisons at single-cell resolution remain scarce. To explore this diversity"

First round Response: We thank the reviewer for noticing this. We have corrected the sentence in the revised manuscript at line 33.

Second round reviewer: solved.

First round 7) Clarify the statistical process for normalizing pseudo-bulk samples across libraries and species.

First round Response: We thank the reviewer for the request for clarification. We've used two different approaches to calculate normalized pseudo-bulk expressions in our study. First, for the detection of divergently expressed orthologs and calculation of cell type specificity, the normalized pseudo-bulk expression was computed by normalizing the raw gene counts in each cell by the total counts of that cell, and these normalized counts were then averaged across all cells belonging to the same cell type. These normalized pseudo-bulk expressions were z-score normalized across cell types prior to Mfuzz clustering in the analysis to identify divergently expressed orthologs. Second, for the quantification of transcriptomic distance, we aggregated the raw counts across all cells of the same cell type and performed variance stabilizing transformation using DESeq2 as suggested in the original DESeq2 paper. We have revised our Methods to explicitly describe the different approaches used for different analyses at lines 818-823, 827-829, 835-837, and 890-894.

Second round reviewer: Solved.

First round 8) The results section for the abstract and introduction needs more detail and clarification; the descriptions are very general.

First round author response: We thank the reviewer for this important comment and have revised both the Abstract (at lines 32-52) and the concluding paragraph of the Introduction to provide a more specific, results-focused summary of our main findings (at lines 82-83 and 87-95). Specifically, the Abstract now states the scope of the dataset (>200,000 root cells across five Brassicaceae species

under control, NaCl, and ABA) and summarizes novel findings, including the limited conservation of canonical Arabidopsis marker specificity across Brassicaceae, the mapping of cortex subpopulations to discrete cortical layers across species by in situ hybridization, and evidence for cortex functional specialization via preferential localization of flavonoid biosynthesis programs to the inner cortex layer. We also clarified the comparative outcomes on stress responses by explicitly stating that cell-type contributions differ across species and treatments and by summarizing the observed evolutionary turnover in responsiveness. We kept a concise but specific summary of the polyploid result in *C. sativa*. In parallel, the final paragraph of the Introduction was rewritten as a clearer preview of the study's main outputs, including marker conservation/refinement, phylogeny-aware shifts in ortholog expression profiles, cortex layer mapping and specialization, lineage- and treatment-dependent cell-type stress responses with turnover patterns, and polyploid subgenome/homeolog diversification.
Second round reviewer: solved.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

I would like to thank the authors for their thoughtful, clear, and detailed responses to my questions and suggestions. All of my concerns have been addressed in the revised manuscript and I have no further suggestions. It was a genuine pleasure to review this work.

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)

I thank the authors for their efforts in addressing all the questions and comments. In particular, they clarified the methodologies and the results presented in Figure 5, which makes the manuscript much easier to understand. Although the authors did not perform DAP-seq experiments, they provided substantial additional analyses, including new analyses based on publicly available data. Lastly, the revised abstract is much clearer than the previous version, and the novelty and impact of the study are now better articulated. This study will serve as a strong and timely contribution and a valuable resource for the community to better understand divergent stress responses across different cell types during Brassica evolution.

(Remarks on code availability)

I tried the website again this time, and it can run smoothly.

credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

RESPONSE TO REVIEWER COMMENTS

We sincerely thank all the reviewers for their time and thorough assessment of our manuscript. We are grateful for the constructive feedback and insightful suggestions, which we believe have significantly strengthened the clarity and biological interpretation of our study. We also appreciate the recognition of the value of our cross-species single-cell transcriptome study. We have carefully revised our manuscript to address all the comments as detailed in the point-by-point responses below. Reviewers' comments are shown in **black**, and our responses are provided in **red**.

Reviewer #1 (Remarks to the Author):

This study presents a comprehensive single-cell transcriptomic atlas of root cells from five Brassicaceae species—*Arabidopsis thaliana*, *Sisymbrium irio*, *Eutrema salsugineum*, *Schrenkiella parvula*, and *Camelina sativa*—under control, ABA, and salt stress conditions. Using 10X Genomics scRNA-seq, the authors generated over 174,000 root cell transcriptomes and developed a cross-species annotation framework (for 4 of the 5 species). They found that nearly half of *Arabidopsis* cell-type marker genes lacked conserved expression across species, prompting the identification of a new set of conserved Brassicaceae markers. The study revealed lineage-specific anatomical and transcriptional variation, especially in cortex subpopulations, and showed that stress responses are highly cell-type and species-specific. In *Camelina* sub-genomes seemed to contribute equally to stress responses, and divergence among homeologs was largely uncoupled from coding sequence variation. While the study identifies divergent expression patterns, functional validation of these genes is limited, however the validations that were done seem reasonable. I am not expecting the authors to do more validations.

A few details are missing that I hope could be included or expanded upon, however the work is very solid, and the authors are using state of the art methods for analysis- especially the prediction algorithms, I have no major notes on types of analysis. Because of the nature of stress (reduction in many types of cell-type expression, see Jean-Baptiste heat shock single-cell), a major concern would be that more abundantly sampled cell types would have the strongest signal simply due to power. However, the authors do their best in compensating for that in several ways. The *Camelina* work could use more expansion and feels very much like an afterthought or even a completely different paper, loosely related to the expansive, integrated analysis in the rest of the paper. After addressing the relatively straight forward requests below I can only suggest publication so strongly.

- Quiescent center cells were underrepresented (really, completely absent), limiting insights into stem cell niche dynamics. I would love to hear more about why this is, most other studies in *Arabidopsis* (even those not using root tips) have at least some QC present. The only mention is just that it is under-sampled, but with 100,000+ cells I'm not sure this is the only reason.

Response: We thank the reviewer for this thoughtful and important observation. We have carefully re-examined the absence of QC cells in our dataset. To determine whether the absence of QC cells is due to excessive stringency of our analyses, we relaxed cell filtering thresholds to keep all cells, and mapped unfiltered cells to Shahan's reference Arabidopsis atlas¹. The projection showed that there were no QC-like cells present in our dataset, see **Response Fig. 1**. Combined with the observation that we did not detect the expression of QC markers in our dataset as shown in **Fig. 1d and Supplementary Fig. 8**, we confirmed that the QC cells are truly absent in our dataset.

We believe that the apparent absence of QC cells in our dataset reflects the spatial scale and experimental design of our sampling and is largely due to under-sampling. In our study, we harvested the lower **0.7-1 cm** from the root tip across species for scRNA-seq, in contrast to the **~0.5 cm** region typically used in prior studies. Because the QC occupies a very small domain at the apex and comprises only a handful of cells per root, extending the sampling size dramatically increases the total cell count without increasing the QC count. Our harvesting strategy prioritized maximizing the detection of stress-responsive signals rather than focusing on shorter sections enriched primarily for stem cells at the root apex. This effectively diluted the proportion of QC cells in our samples. A comparative meta-analysis of existing Arabidopsis root atlas confirms that successful QC detection is largely linked to sampling a smaller region (~0.5 cm) of the root tip, see **Response Table 1**.

More importantly, the scope of our study is to map cell type-specific stress responses and the absence of QC cells does not compromise the validity of our findings. Our analytical framework treats cell types as independent modules, and the statistical power and accuracy of all our analyses are mathematically independent of the presence or absence of the QC cells. Thus, our dataset remains robust for characterizing the transcriptomic landscape of cell types that were successfully captured and resolved.

[Figure Redacted]

Response Fig. 1. Projection of Arabidopsis cells from current study to Shahan *et al.* (a) The Arabidopsis reference atlas from Shahan *et al.*¹. The projection of unfiltered Arabidopsis cells from control condition (b) and unfiltered Arabidopsis cells from treated conditions (c) onto the reference atlas. Right circles highlight regions where QC cells are expected to locate at.

Study	Plant Age	Sampling Strategy	Total Cells	QC Cells Detected?
Denyer et al. (2019)	6 days	Tip (~1.0 cm)	4,727	Yes
Ryu et al. (2019)	5-7 days	Tip (0.5 cm)	7,522	Yes
Zhang et al. (2019)	10 days	Tip (0.5 cm)	7,695	Yes
Wendrich et al. (2020)	6 days	Meristem (Tip)	5,145	Yes
Shahan et al. (2022)	5 days	Tip (0.5 cm)	110,427	Yes
Jean-Baptiste et al. (2019)	7 days	Whole Root (~2 cm)	3,121	No
Shulse et al. (2019)	5-7 days	Whole Root	12,198	No
This study	4-6 days	Lower 1/3 (~1.0 cm)	>200,000 (from five species)	No

Response Table 1. Comparative summary of Arabidopsis root single-cell datasets and sampling parameters.

- The analysis of sub-genome dominance in *Camelina* is largely descriptive and lacks mechanistic depth. There is very little description of what has to be a challenging problem, dealing with hexaploidy genes, was the alignment rate drastically less? Were there particular alignment strategies? Etc. The methods do not have any of these types of details either. Presumably these types of challenges are why *Camelina* was singled-out and not included or attempted to be included in the cross-species analysis, but this is not addressed as far as I read.

Response: We thank the reviewer for highlighting the need for clarification in our methodology regarding the *C. sativa* analysis. The exclusion of *C. sativa* from the integrated cross-species comparisons was a strategic choice to preserve sub-genome resolution in this species and also simplify cross-species comparisons.

As an allohexaploid species, *C. sativa* contains three sub-genomes (G1, G2, G3) and sequence divergence among the sub-genomes is sufficiently high to allow standard aligners (STAR/Cell Ranger) to map reads uniquely to specific homologs, as demonstrated in previous studies^{2,3}. Therefore, we've used the same standard alignment strategy for all the species used in our study. The mapping rate for *C. sativa* (~93%) was highly comparable to that for the diploid species, like *Arabidopsis* (~97%), see **Supplementary Data 1**.

The challenge in integrating *C. sativa* with the rest of the species mainly lies on orthology constraints. Cross-species comparisons in our study relies on 1-to-1 orthologs, and including the hexaploid *C. sativa* introduces a complex 1-to-3 orthology relationship. Because of this, we would need to either (a) aggregate homeologs by averaging their expression, (b) use a representative sub-genome for data integration with other diploid species, or (c) separate the sub-genomes. However, both (a) and (b) would mask the sub-genome resolution in *C. sativa*, and (c) splits one sequenced *C. sativa* cell into three virtual cells with each of them only expressing a fragment of the whole cell signature. By analyzing the hexaploid transcriptomes as a holistic and independent unit, we preserve the full transcriptional profile required to robustly define cell identity, which then allows us to retrospectively investigate sub-genome contributions as we have done in **Fig. 7**.

To address the reviewer's concern here, we have added clarification for why we singled out *C. sativa* in the revised manuscript at lines 105-107 and 114-117.

- The overall UMIs are on the lower side, could you please provide knee plots for these experiments in the supplement.

Response: We thank the reviewer for this suggestion. We have now provided barcode knee plots for all the scRNA-seq experiments in **Supplementary Fig. 3**. These plots show a clear inflection point separating high-confidence cell-associated barcodes from background ones, indicating that the observed UMI distributions reflect genuine cell-associated transcript capture rather than an excess of low-complexity or ambient RNA barcodes. Despite the modest UMI depth, our data

robustly resolved all major root cell types across species and recapitulated established marker gene expression, consistent with prior root scRNA-seq studies.

- There is a period missing on line 2 of the abstract (line33).

Response: We thank the reviewer for noticing this. We have added the period in the revised manuscript at line 33.

- Perhaps you could further expand on line 204-205, when you say “This discrepancy may be due to the technical differences between the two complimentary RNA detection methods”

Response: We thank the reviewer for this suggestion. The difference between the scRNA-seq and HCR results likely stems from two fundamental methodological distinctions. First, HCR often offers better *in situ* sensitivity due to its signal amplification mechanism. Combined with its preservation of tissue architecture, this allows for the detection of weak and spatially restricted signals. These subtle, localized expression patterns are often subject to “dropout” or dilution in droplet-based scRNA-seq libraries, particularly if the transcripts are in low-abundance or present only in a small cell population. Second, scRNA-seq requires enzymatic digestion to isolate protoplasts, which can induce rapid transcriptional turnover for some transcripts compared to the immediate fixation used in HCR.

In case of the discrepancy in HCR and scRNA-seq results for *GSTU25*, we believe that this is mostly due to the higher sensitivity of HCR in detecting gene expression in spatially restricted regions. We have expanded our explanation on lines 222-224 to explicitly detail the potential cause of this discrepancy.

- I also wonder if some of the cell type differences are simply like structural/lifestyle differences, where the actual physical differences of the roots leads to different penetration etc. of the media stresses.

Response: We appreciate the reviewer raising this important point regarding the anatomical/structural differences among species. While we agree that anatomical and structural differences among species can influence how environmental stresses are perceived spatially within the root, we believe the transcriptional differences reported here are not simply structural or passive consequences of physical stress penetration.

First, we observed substantial changes in cell type specificity of orthologs across species (**Supplementary Fig. 12**), and have identified many instances of “spatial rewiring” where orthologs are expressed in different cell types across species, including *NIP5;1* (**Fig. 3c and 3e**), *GSTU25* (**Fig. 3d and Fig. 3f**), alongside with many other divergently expressed orthologs listed in **Supplementary Data 4**. If cell type differences were merely passive results of stress penetration, we would expect most orthologs to maintain conserved cell type expression profiles across species, with only their expression levels changing based on how much stress reaches

them. Instead, our observations of the spatial expression divergence of orthologs contradict the passive perception argument and point to a real evolutionary regulatory divergence.

Second, we found functional specification of cortex layers in *S. irio* and *S. parvula* where there is an additional cortex layer compared to Arabidopsis (**Fig. 4d, 4e, 4f, and Supplementary Fig. 15**). The inner cortex layer in both species has evolved to express a distinct transcript population (**Supplementary Fig. 16**). For example, the flavonoid biosynthesis pathway is largely biasedly expressed in the inner cortex layer in both species (**Fig. 4f**). This suggests that having an additional cortex layer serves beyond creating a passive physical barrier, but also serves as a metabolically specialized cell layer.

Third, we observed robust and statistically significant transcriptional remodeling in deeper cell layers across all species and the magnitude of these transcriptional responsiveness does not necessarily follow a centripetal gradient of decay from the root surface (**Fig. 5b and Supplementary Fig. 23**). The stress responsive genes in the deeper cell layers, like the stele, often exhibited fold-changes comparable to those in outer cell layers, like epidermis (**Supplementary Fig. 20-22**). These indicated that the 3h ABA and 24h NaCl treatments successfully permeated through all cell layers and triggered a systemic transcriptional response throughout the radial axis, thereby largely minimizing the limited accessibility as a confounding factor.

Fourth, the observation of “stress preparedness” - defined as the constitutively higher or lower basal expression of stress-responsive genes in the extremophytes (**cluster 0 and 6 in Fig. 6d**) - across all cell types also suggested that their transcriptional divergence is “genetically fixed” into their basal regulatory networks. Since this transcriptional divergence is present in control conditions when no stress treatment is applied, it is fundamentally independent of the kinetics of stress media penetration or stress accessibility.

Finally, the root structure itself is an active physiological response, not a static feature. For example, Li *et al.*⁴ found that *S. parvula* actively upregulates MYB transcription factors and suberin biosynthesis genes within 3 hours of salt treatment to extend the suberized zone in the endodermis, while Arabidopsis fails to do this under similar conditions. Similarly, Tran *et al.*⁵ revealed that *S. parvula* actively expands its xylem vessels to maintain water flow and actively maintain its growth under stress. These studies demonstrated that the root’s structural response is plastic driven by the diverse transcriptional networks we aimed to characterize.

Therefore, we believe that the differences we observed in cell type stress responses across species represent true cell type-specific and lineage-specific adaptations of the transcriptional machinery, rather than passive outcomes of differential media penetration or stress accessibility.

Reviewer #2 (Remarks to the Author):

In this interesting manuscript, Wang et al., explored the diversity of root cell types and their expression patterns across the Brassicaceae family under salt and ABA stresses. The authors performed a multi-species single-cell comparative transcriptomics of the root, first to identify conserved and non-conserved cell type gene markers. They found that almost half of Arabidopsis gene markers are not conserved; however, through evolutionary and comparative analyses, they defined a set of conserved markers. Second, they identified transcriptomic evidence that supports the specialization of cortical cell subpopulations in *S. parvula* and *S. sirio*. Third, they observed a divergence in stress responses among species, especially in response to salt stress. Fourth, one of the most interesting contributions in this manuscript is the stress response between orthologs, trying to link the transcriptomic shifts to phylogenetic relationships, confirming the cell type contributions to the stress-prepared transcriptomes in extremophytes. Finally, they used the *Camelina Sativa* single-cell RNA-seq data to explore the effects of polyploidy on cell-type specificity.

Overall, this manuscript is of great significance to the root biology community, serving as a perfect complement to the root transcriptomes of numerous plants at single-cell resolution. This enhancement will undoubtedly provide valuable resources to the plant community. While this manuscript has the potential to be published in the journal, it requires further clarification and restructuring. I have attached my comments below for your reference.

Major:

1) One of the most important results of this manuscript is that the authors identified half of the markers for Arabidopsis as not conserved across the Brassicaceae family. My concern here is that this result is a reflection of the low gene expression because of the library size or the plant species. How can we be confident that these “lost” of expression is not noise across the gene expression between species? Could the authors clarify in the main text and methods the normalization of the expression per library size, plant species, and batch effects? Moreover, I was wondering what happened with the other half of the Arabidopsis markers. Are they entirely conserved across the Brassicaceae family?

Response: We thank the reviewer for this important concern and agree that it is essential to distinguish true evolutionary divergence from technical noise.

In our analysis, cell-type markers were identified using Seurat’s FindMarkers within each species, where markers were defined as genes showing consistent and cell-type-enriched expression relative to their expression in other cell types in the same species. Thus, by definition, markers represent genes with sufficient expression levels to be reliably detected and enriched within a given dataset. To examine the expression of these cell-type markers across species, we used a cell-type pseudo-bulk strategy, in which raw gene counts in each cell were normalized by the total counts of that cell and the resulting normalized values were averaged across all cells of the same cell type. This approach captures the average per-cell transcriptional state while

minimizing effects of cell number and library size differences. We have clarified these details in the Methods at lines 818-823.

When assessing conservation across species, we observed that genes classified as “non-conserved” were not universally lowly expressed. Instead, these genes often showed substantially detectable expression and clear marker behavior in one or more species, while being lowly expressed in other species, where they therefore could not be identified as valid markers (**Supplementary Fig. 9**). Importantly, orthologs of these genes in other Brassicaceae species frequently retained robust expression and cell-type specificity. Thus, “non-conserved” refers to species-specific attenuation or loss of marker-level expression, rather than technical dropout or global loss of expression.

With respect to the remaining half of Arabidopsis markers, these genes retained consistent expression levels and cell-type specificity across all Brassicaceae species examined and were therefore classified as conserved markers (**Supplementary Fig. 9**). This list represents the most robustly conserved root cell-type markers currently known for Brassicaceae species, although this list will likely be refined and expanded as additional root single-cell transcriptome datasets become available for other members of the family. We have clarified these details in the Results at lines 155-160.

2) I found interesting the cortex specialization in *S. parvula* and *S. Sirio*. Which are the top-expressed genes in each cortex layer? What is the transcriptomic signature or co-expressed genes for that extra cortical layer? More importantly, what happens if you extend this comparative analysis of cortical cells using available data for tomato and *Medicago truncatula* (two species with two layers of cortical cells)? Are the same genes involved? I suggest adding at least these cortical cell data of the root to extend the results and generalize about this specialization process in roots.

Response: We thank the reviewer for raising these important questions. We have performed the following analyses to address these questions.

To identify the transcriptomic signatures of the inner and outer cortex layers, we compared the transcriptomes of different cell layers (with cortex divided into inner and outer layers) in *S. irio*, and identified the genes co-expressed for the different cortex layers (**Supplementary Fig. 17**). We found that genes involved in water transport and environmental sensing were highly expressed in both inner and outer cortex layers compared to other cell types (**Supplementary Fig. 17**). Genes preferentially expressed in the inner cortex were mostly associated with secondary metabolism (particularly flavonoid biosynthesis), cell-to-cell communication via plasmodesmata, and jasmonic acid and ABA signaling. In contrast, the outer cortex prioritized the expression of genes associated with hypoxia sensing, thermal stress response and gibberellin signaling pathways (**Supplementary Fig. 17**).

To investigate the cortex specialization in other species, we extended our analysis to other species with more than one cortex layer in roots and corresponding root single cell transcriptome data, including rice⁶ and Medicago⁷. We did not include tomato as we were not able to locate a sufficiently-annotated dataset for this species. We extracted cortex cells from rice and Medicago and applied the same within-species clustering approach to these cells to maximize the sensitivity for detecting internal subpopulation structures. We identified two major subpopulations in the rice cortex cells, corresponding to the elongation and maturation zones based on the developmental zone annotation (**Supplementary Fig. 18**). Notably, the rice cortex cells exhibited a single, linear progression from elongation to maturation, in clear contrast to the branching paths we observed for cortex cells in *S. irio* and *S. parvula* in our study. This lack of distinct branching trajectories suggests that, at this resolution, rice cortex cells maintain a relatively unified transcriptional identity as they transition toward their terminal functional states, with developmental stage being the primary driver of transcriptomic variance rather than radial cell-layer specialization. We also detected the subpopulation structure for the Medicago cortex cells (**Supplementary Fig. 18**), however, we were unable to clearly determine the biological relationship of the subpopulations due to the lack of reliable developmental annotation for the Medicago dataset, preventing a definitive comparison of its trajectory direction. It is important to note that because we primarily examined cells from stress-neutral conditions, we cannot rule out the possibility that cortex specification in these species is condition-dependent or inducible upon environmental changes.

To further investigate whether this transcriptomic uniformity reflects a lack of functional specialization, we examined the expression of flavonoid biosynthesis genes, which are differentially partitioned between different cortex layers in our data. Our analysis demonstrated that these genes are either undetected, including CHS which catalyzes the first committed step in the flavonoid biosynthetic pathway, or universally expressed across all cortex cells, rather than being limited to a specific subset of cortex cells (**Supplementary Fig. 18**). This absence of partitioned metabolic signatures reinforces our subpopulation analysis results and suggests that, even if cortex specification does exist in these species, the flavonoid pathway is not involved in that specialization. We have added the results of these additional analyses to the revised manuscript at lines 256-266 and 278-285 (the corresponding method is added at lines 957-963).

3) I'm concerned about the restructuring of the paper. In line 93, the authors described the number of cells for each plant species (5 root species), but surprisingly, in Fig. 1, only 4 plant species appear. What happened with *Camelina sativa*? Are there some contradictions. The authors also mentioned that they generated between 36,000 and 50,000 root cell transcriptomes for the five plant species. However, as shown in Fig. Supp 1, *Camelina sativa* has 27,442 cells. I can imagine that could be related to the restructuring of the manuscript, because at the end of the manuscript, the last section (involving Fig. 7) is difficult to follow, with likely a disconnection of the main story. Starting again with a UMAP, and those results are described at the beginning.

Response: We thank the reviewer for raising this important point regarding the clarity and organization. We agree that the current presentation could be difficult to delineate and that the handling of the *C. sativa* dataset requires additional clarifications.

As described earlier in response to Reviewer 1, because of the challenge in integrating polyploid species with diploid species, we intentionally divided our manuscript into two related but distinct sections: 1) a comparative single-cell analysis of four diploid Brassicaceae species, and 2) a separate application of the same framework to the polyploid crop *C. sativa*. This is because all our cross-species comparisons rely on 1-to-1 orthologs, and including the hexaploid *C. sativa* introduces a complex 1-to-3 orthology relationship. Therefore, we focused **Fig. 1-6** exclusively on the comparisons among the four diploid species. *C. sativa* is introduced later as a dedicated case study to evaluate cell-type annotation transfer and sub-genome contributions in a polyploid context as shown in **Fig. 7**. We have revised the manuscript to explicitly explain the layout of our manuscript and frame *C. sativa* as a separate extension of the analytical framework, rather than a continuation of the main diploid comparison.

Regarding the cell numbers, the statement that “36,000 to 50,000 root cell transcriptomes were generated for each species” refers specifically to the four diploid species. We’ve separately profiled over 27,000 cells for *C. sativa*, which is lower due to differences in protoplast recovery and sequencing yield. We have clarified this in the revised manuscript at lines 105-107 and 114-117.

4) Figure 5 is challenging to comprehend. Panel A is too small. Panel B illustrates the transcriptome response (I can envision differential gene expression) for each cell type and treatment. Here, it’s crucial to quantify the number or percentage of cells responding to each stress in each plant species. In this way, the authors can identify the cell types with the most specific response. Please clarify panel E.

Response: We thank the reviewer for this helpful feedback and agree that **Fig. 5** requires clearer presentation and detailed explanations. We have enlarged **Fig. 5a** for better readability.

Fig. 5b does not represent the number or percentage of cells responding to the treatments, but rather the degree of transcriptomic separability between control and treated cells within each cell type in each species, as quantified by the Augur framework⁸. Augur measures the degree to which treated cells can be distinguished from untreated cells based on their overall transcriptomic profiles, yielding an AUC score for each cell type. Cell types in which treated and control cells are easily separable (reflected as having a high AUC score) are interpreted as having stronger transcriptional responses or shifts, whereas cell types in which treated and control cells remain similar (reflected as having a low AUC score) are interpreted as having weaker responses or being unchanged upon exposure to stress. The results reflect the magnitude of transcriptomic change for a cell type in response to the treatment, rather than the number or proportion of responding cells. Because of the high sparsity in scRNA-seq data and also because we’ve used “pseudo-bulk” differential expression analysis to identify differentially expressed

genes as it has been shown to achieve the highest fidelity^{9,10}, it is currently impossible and not reliable to directly quantify the number or the percentage of cells responding to stress. However, we have provided the number of differentially expressed genes per cell type and treatment in each species in **Supplementary Fig. 23 and Supplementary Data 6**, which offers a complementary and more intuitive measure of cell type-specific responsiveness.

Fig. 5e aims to determine which cell type shows the largest transcriptomic divergence across species and the impact of treatments on the cross-species transcriptomic divergence. Each data point represents the transcriptomic divergence between six pairwise species comparisons (shown as the x-axis in **Fig. 5f**) under a given condition. On the PCA space shown in **Fig. 5e**, columella cells from different conditions form a unique cluster that is more separated from the rest of the cell types on PC1. This is due to columella cells exhibiting the largest transcriptomic divergence across species (**Fig. 5f and Supplementary Fig. 27**), suggesting that the columella cells from these species are highly dissimilar to each other, aligned to what has been reported for grass family species¹¹. We have also added a supplementary figure (**Supplementary Fig. 26**) to quantitatively measure the distances between treated and control samples.

5) The title suggests that the study focuses on the evolutionary diversity and stress response in roots. However, this study is in the Brassicaceae family.

Response: We thank the reviewer for raising this point. We agree that the study is focused on species within the Brassicaceae family, and our intent with the title was not to imply evolutionary comparison across all plant lineages. To avoid any possible ambiguity, we have revised the title to explicitly reference the Brassicaceae family at line 1.

Minor:

6) Line 33. typo: "yet cross-species comparisons at single-cell resolution remain scarce. To explore this diversity"

Response: We thank the reviewer for noticing this. We have corrected the sentence in the revised manuscript at line 33.

7) Clarify the statistical process for normalizing pseudo-bulk samples across libraries and species.

Response: We thank the reviewer for the request for clarification. We've used two different approaches to calculate normalized pseudo-bulk expressions in our study. First, for the detection of divergently expressed orthologs and calculation of cell type specificity, the normalized pseudo-bulk expression was computed by normalizing the raw gene counts in each cell by the total counts of that cell, and these normalized counts were then averaged across all cells belonging to the same cell type. These normalized pseudo-bulk expressions were z-score normalized across cell types prior to Mfuzz clustering in the analysis to identify divergently expressed orthologs. Second, for the quantification of transcriptomic distance, we aggregated the

raw counts across all cells of the same cell type and performed variance stabilizing transformation using DESeq2 as suggested in the original DESeq2 paper. We have revised our Methods to explicitly describe the different approaches used for different analyses at lines 818-823, 827-829, 835-837, and 890-894.

8) The results section for the abstract and introduction needs more detail and clarification; the descriptions are very general.

Response: We thank the reviewer for this important comment and have revised both the Abstract (at lines 32-52) and the concluding paragraph of the Introduction to provide a more specific, results-focused summary of our main findings (at lines 82-83 and 87-95). Specifically, the Abstract now states the scope of the dataset (>200,000 root cells across five Brassicaceae species under control, NaCl, and ABA) and summarizes novel findings, including the limited conservation of canonical Arabidopsis marker specificity across Brassicaceae, the mapping of cortex subpopulations to discrete cortical layers across species by in situ hybridization, and evidence for cortex functional specialization via preferential localization of flavonoid biosynthesis programs to the inner cortex layer. We also clarified the comparative outcomes on stress responses by explicitly stating that cell-type contributions differ across species and treatments and by summarizing the observed evolutionary turnover in responsiveness. We kept a concise but specific summary of the polyploid result in *C. sativa*. In parallel, the final paragraph of the Introduction was rewritten as a clearer preview of the study's main outputs, including marker conservation/refinement, phylogeny-aware shifts in ortholog expression profiles, cortex layer mapping and specialization, lineage- and treatment-dependent cell-type stress responses with turnover patterns, and polyploid subgenome/homeolog diversification.

Reviewer #3 (Remarks to the Author):

In this manuscript, Wang, Ryu, Dinneny, and colleagues used scRNA-seq to create a cross-species transcriptomic atlas of Brassicaceae roots treated with abscisic acid and NaCl. The authors found that the expression of a majority of Arabidopsis thaliana root cell type markers are not conserved in the other profiled species. Thus, validation of newly identified markers via HCR produced a toolkit of conserved markers and revealed orthologs with lineage-divergent expression profiles. Taking advantage of their atlas resource, the authors investigated how different cell types respond to stress between species and across treatment types.

Overall, I believe that the computational framework established for cross-species and pan-treatment analysis of scRNA-seq data will surely be a useful resource for the plant biology community. However, to me, the significance of the biological insights in the manuscript are less clear. I will expand on these thoughts and provide additional comments below.

1. Although the authors undertook a comprehensive annotation of root cell types, there is no mention of the different developmental zones of the root, save for the labels in Figure 3, panels E

and F. I assume that the data captures primary root tips through at least the early differentiation zone, but there is no indication of how much of the root was harvested for scRNA-seq for each species (lines 551-552). Further, although different parts of the root (e.g., meristem versus elongation and/or differentiation zones) are shown in Figure 2 to correspond to expression of the different marker genes, it is unclear if these differences in spatial localization match the scRNA-seq data.

Response: We thank the reviewer for raising this important point. We've harvested the top one third of the root tips (0.7-1 cm) from seedlings that had grown about 2-3 cm long for each species. The harvested root tissues included root cap, meristematic, elongation, and differentiation zones. We added the corresponding developmental annotations to our dataset (lines 310-312), now present as **Supplementary Fig. 6**. Based on the developmental annotations, *NIP5;1* is expressed primarily in the differentiation zone (the terminal) population of endodermal cells in *S. parvula*, which was consistent with our HCR results for *NIP5;1* in *S. parvula*. The expression of *GSTU25* in *E. salsugineum* and *S. parvula* was observed to start in the endodermal cells in the meristematic zone, and present until the early elongation zone. The HCR results for *GSTU25* in these two species are also largely in agreement with these observations. We have revised our manuscript to include these results at lines 206, 216-217, and 637-639.

Can the authors please clarify how much of the root was harvested for each species, what developmental zones were captured in the scRNA-seq data, and how they determined in which developmental zone the putative marker genes were expected to be expressed in the HCR experiments? Additionally, can the authors address if markers expressed in specific developmental zone(s) will be useful for understanding stress responses (e.g., is it an open question if cells more responsive to stress depending on their differentiation state)?

Response: We thank the reviewer for raising this important request for clarification. For all species, we've harvested approximately the top one third of the root tips (0.7-1 cm) from seedlings that were about 2-3 cm long. This included all the major developmental zones of roots, including the meristematic, elongation, and differentiation zones.

In our original manuscript, we did not explicitly analyze marker genes or stress responses across developmental zones. In the revised manuscript, we've added the corresponding developmental zone annotations to our dataset (**Supplementary Fig. 6**) and identified markers for each developmental zone (**Table S3**). We further applied the same pseudo-bulk differential expression analysis to identify ABA- and NaCl-responsive genes within each developmental zone. This analysis revealed differences in the number of differentially expressed genes among different developmental zones, and these differences vary among different species (**Supplementary Fig. 24, Supplementary Data 6**). Furthermore, we found that a substantial proportion of stress-responsive genes were specific to individual developmental zones, indicating that each zone exhibits distinct transcriptional responses to the treatments (**Supplementary Fig. 24**).

Together, these results demonstrate that developmental stage is an important dimension of stress-responsive transcriptional variation. We have added these analyses to our revised manuscript at lines 310-312 and 328-330 to highlight that our dataset enables examination of whether stress responses are preferentially associated with particular cell types and developmental stages, and this remains an important direction for future investigation.

2. Figure 4; Lines 211-253; Supp Fig 13. To me, this section seems out of place with the stress-response focus of the rest of the manuscript. Can the authors provide context for the significance of a cortex layer that specializes in flavonoid metabolism? Are flavonoids important for stress responses or plant resilience?

Response: We thank the reviewer for this important comment and agree that the relevance of this section to the broader stress-response framework was not sufficiently articulated.

Flavonoids play well-established roles in plant responses to abiotic stress, including functioning as antioxidants¹², modulators of reactive oxygen species¹³, and regulators of hormone signaling and auxin transport¹⁴. Accumulation of flavonoids has been linked to enhanced tolerance to salinity, drought, and oxidative stress¹⁵, and the spatial restriction of flavonoid biosynthesis to specific tissues and cell layers has been proposed to modulate these protective functions¹⁴. For example, flavonoids accumulate in cortex cells of the root elongation zone in *Arabidopsis* and function as endogenous negative regulators of basipetal auxin transport, therefore influencing root growth and lateral root development under different environmental contexts¹⁶. Our finding that an additional cortex layer is present in *S. irio* and *S. parvula* and this layer is transcriptionally specialized for flavonoid biosynthesis therefore suggests a lineage-specific anatomical partitioning of a growth and stress-relevant metabolic pathway.

We have revised the manuscript at lines 278-285 and 582-586 to more explicitly discuss the role of the cortex-layer specialization by highlighting the established roles of flavonoids in regulating plant growth and responses to environmental stresses and framing the emergence of an inner cortex layer with enriched flavonoid biosynthesis as a potential evolutionary innovation that may contribute to plant growth and stress responses in these lineages.

3. Figure 5, lines 325-338. The identification of cell type transcriptional profiles that diverge across species and treatments is interesting but the biological context of these results is somewhat lacking. For example, the authors note that procambium cells showed the most increases in transcriptomic divergence between species, particularly in response to ABA (lines 336-337). Is the identification of procambium particularly interesting compared to other cell types? Is procambium function known to differ across species?

Response: We thank the reviewer for raising this important point and agree that the biological context of this result requires clearer explanations.

Procambium cells are known to function as the primary meristematic tissue providing precursors that differentiate into vascular tissues¹⁷. They occupy a developmental niche that integrates

growth, hormone signaling, and long-distance transport¹⁸. Procambium activity has been shown to be sensitive to stress and hormone signaling, including ABA, which can modulate vascular differentiation and transport-related gene expression¹⁹. As such, they serve as a key interface between developmental patterning and environmental response.

While the fundamental developmental role of the procambium is conserved, the spatial organization of vascular tissues varies significantly between lineages, suggesting a link to diverged functions of the procambium²⁰, although direct experimental evidence for this remains limited. The observation that procambium cells show particularly strong transcriptomic divergence across species in response to ABA suggests that evolutionary changes in stress signaling may preferentially affect tissues involved in coordinating systematic growth and transport, rather than only terminally differentiated cells.

We do not interpret this result as evidence that procambium uniquely determines stress adaptation. Rather, it highlights a developmental cell type whose stress-responsive transcriptional programs may have undergone significant lineage-specific rewiring. We have revised our manuscript to provide this context at lines 384-388.

4. Figure 6, lines 392 – 419. The authors investigated whether the previously reported stress-prepared state at the organ level in extremophytes is due to specific cell types. In response to salt treatment, the authors found that outer layer cells (LRC and atrichoblasts) contribute to high constitutive gene expression in the profiled extremophyte species. They conclude that this result suggests a tendency for some cell types to play a more substantial role than others in shaping the stress-prepared signal. Could the authors address if this response could be due to the fact that the outer layer of cells is directly exposed to the environment and to salt treatment in the experiment? Couldn't it therefore be expected that they would respond differently than other cell types?

Response: We thank the reviewer for this thoughtful comment and agree that the physical exposure of outer root layers to the external environment is an important factor to consider when interpreting these results. Indeed, LRC and atricoblast cells are among the first cell types to encounter salt treatment, and this positional context likely contributes to their more prominent stress-prepared transcriptomes. However, surface exposure alone does not fully explain our observations. Other surface cell types, like columella cells and trichoblasts, did not exhibit similarly elevated contributions to the organ-level stress-prepared state (**Fig. 6d**). Moreover, in response to ABA, more internal cell types, including cortex and pericycle cells, showed stronger contributions to the stress-prepared signal than the surface cell types (**Fig. 6d**).

Together, these results suggest that the contribution of individual cell types to organ-level stress preparedness is likely shaped by a combination of their spatial position within the tissue, the nature of the stress, and the cell type-specific capacities for stress perception and signal transduction. We have revised the manuscript at lines 478-485 to clarify that these patterns likely

reflect an interplay between anatomical position and cell type-specific physiological roles, rather than a simple consequence of surface exposure.

5. Figure 7. The authors demonstrate that their cross-species cell type annotation strategy enables successful annotation of scRNA-seq data from *C. sativa*. This analysis revealed a lack of large-scale sub-genome dominance in stress responses (Lines 450-453). I don't fully understand the implications of this conclusion. Could the authors clarify how sub-genome dominance (or lack thereof) in stress responses is useful for or a hindrance to plant resilience?

Response: We thank the reviewer for this important question and agree that the biological implications of this result required clearer explanation.

In polyploid species such as *C. sativa*, sub-genome dominance refers to the preferential contribution of one ancestral genome to gene expression or regulatory responses. If stress responses were dominated by a single sub-genome, this would suggest that adaptive transcriptional programs are largely controlled by one genomic compartment, potentially limiting the functional contribution of the other sub-genomes.

Our findings that stress-responsive transcriptional changes are uncoupled from basal sub-genome dominance and are not strongly biased toward a single sub-genome indicates that all three sub-genomes contribute comparably to stress responses. This distributed contribution may increase regulatory redundancy and buffering capacity, allowing stress-responsive pathways to be maintained even if individual homeologs diverge or are lost. Such a pattern is consistent with the idea that polyploidy can provide robustness to environmental stress by retaining multiple function copies of stress-relevant genes.

In the meanwhile, we want to point out that we do not infer from these data that the absence of sub-genome stress response dominance directly enhances stress tolerance. Rather, our results show that stress-responsive regulatory programs in *C. sativa* are broadly shared across sub-genomes. We have revised the manuscript at lines 605-615 to clarify this interpretation and to avoid implying a direct causal relationship between sub-genome balance and stress resilience.

6. I would like to better understand how the authors envision that the community will use the new resources developed for these Brassicaceae species. For example, beyond confirming the cross-species annotation, will the newly validated cell type markers enable new downstream experiments?

Response: We thank the reviewer for this important question and agree that the utility of these resources should be made more explicit. Beyond enabling cross-species annotation, the validated cell type markers can be used to guide the design of cell type reporters and *in situ* hybridization probes for spatial and temporal tracking of root cell types in multiple Brassicaceae species, especially in species for which such tools are currently limited. The comparative atlas also provides a reference framework for mapping stress-responsive genes to specific cell types and developmental contexts across species, enabling hypothesis-driven experiments to test lineage-

and cell type-specific stress responses and to prioritize candidate genes for functional studies. We have revised the Discussion at lines 556-560 to clarify these anticipated applications.

Minor

1. On the Shiny app, 'atlas' is misspelled in the title at the top of the page.

Response: We thank the reviewer for pointing this out, and have corrected it on the Shiny page.

2. Supplemental Fig 10: It seems like part of the final sentence in the figure legend was omitted.

Response: We thank the reviewer for noticing this, and have included the full legend for **Supplementary Fig. 10** (now **Supplementary Fig. 12**) in the revised manuscript.

3. The processed data deposited to GEO are currently under embargo. Could the authors confirm if annotated Seurat objects are available for the atlas? If not, I suggest providing them to save time and computational resources for the community.

Response: We thank the reviewer for pointing this out. We have created reviewer access for our [GEO series GSE268881](#) with the following access token: **srmfwuqevnuxtun**. In case the reviewer encounters any issues accessing the GEO page, we have attached below the screenshot (**Response Fig. 2**) of the Seurat objects that were included in our GEO. They are the annotated Seurat datasets that were used in our study, and will be available to the public upon the publication of this manuscript.

Supplementary file	Size	Download	File type/resource
GSE268881_integratedAth.RDS.tar.gz	15.5 Gb	(http)	TAR
GSE268881_integratedAthEsaSirSpa.RDS.tar.gz	41.4 Gb	(http)	TAR
GSE268881_integratedCsa.RDS.tar.gz	44.6 Gb	(http)	TAR
GSE268881_integratedEsa.RDS.tar.gz	7.8 Gb	(http)	TAR
GSE268881_integratedSir.RDS.tar.gz	8.6 Gb	(http)	TAR
GSE268881_integratedSpa.RDS.tar.gz	11.8 Gb	(http)	TAR
GSE268881_RAW.tar		(http)(custom)	TAR (of MTX, TSV, TXT)

[SRA Run Selector](#) [?](#)

Response Fig. 2. Screenshot of content included in GEO repository GSE268881.

Reviewer #4 (Remarks to the Author):

This manuscript presents a multi-species single-cell RNA-seq atlas of Brassicaceae roots under control, ABA, and salt stress conditions. The authors profile ~175k cells from *Arabidopsis*, *Sisymbrium irio*, *Schrenkiella parvula*, *Eutrema salsugineum*, and *Camelina sativa*, identify 10 major root cell types across species, and build a comparative framework based on ~15k 1-to-1 orthologs. They show that nearly half of classical *Arabidopsis* root cell-type markers are not conserved in expression specificity in the other species and define a new set of conserved Brassicaceae markers validated by HCR in situ hybridization.

Using phylogenetically informed analyses, they quantify divergence of cell-type-specific expression among orthologs, highlight lineage-specific shifts and extremophyte vs non-extremophyte signatures (e.g. NIP5;1 and GSTU25), and analyze how ABA and NaCl responses redistribute across cell types and species. They further examine the allohexaploid *C. sativa*, showing partial expression dominance of one sub-genome but largely balanced contributions to stress responses and uncoupling of coding versus expression divergence among homeolog triplets. Finally, they provide an interactive browser and processed data matrices as a community resource.

Overall, this is a substantial and timely resource for comparative plant single-cell biology with broadly interesting biological insights into cell-type evolution and stress adaptation.

Major comments:

Figure 5A: Why did the authors not use the UMAP analysis, but PCA? Please justify the choice of the methods.

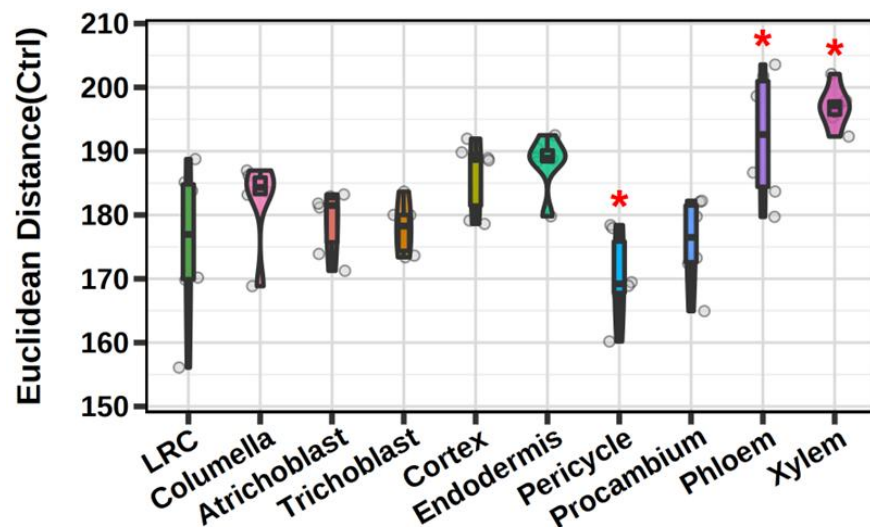
Response: We thank the reviewer for this question. UMAP and PCA serve fundamentally different purposes in single-cell analysis. UMAP is a nonlinear dimensionality-reduction method optimized for visualization of local neighborhood structure, whereas PCA is a linear transformation that preserves global variance structure and relative distances in expression space. Consequently, PCA allows direct interpretation of axes in terms of variance explained and enables unbiased comparison of global transcriptomic relationships across species and conditions. In contrast, UMAP embeddings are sensitive to parameter choices and stochastic initialization and are not metric preserving, making them unsuitable for quantitative comparisons of distances or variance across groups. We therefore used UMAP mostly for visualization and cell-type annotation at the single-cell level, and PCA for comparative analyses of pseudo-bulk transcriptomes where preservation of global structure and interpretability were essential. For Figure 5A, because our goal here was to quantitatively compare transcriptomic similarity and divergence among cell types, species, and treatments rather than to visualize cellular neighborhoods, we used PCA on pseudo-bulk cell-type transcriptomes. We have revised the Methods at lines 742-749 to explicitly state this rationale.

How come the different cell type from the same species show a smaller distance than the same cell types from the different species? I would expect at least some cell types to show a highly conserved expression, and consequently, a shorter distance within the PCA. This makes me question the validity of the method, and whether the observation that the two extremophytes are functionally related. This is also suggested in Figure 5C, where the *S. parvula* and *S. irio* are now more related (due to phylogeny).

Response: We thank the reviewer for pointing this out. It is indeed surprising to see that transcriptomes are grouped primarily by species but not by cell types as shown in **Fig 5a**. However, similar observations have been noted in multiple previous studies that samples often

cluster primarily by species rather than by cell types when comparing single cell data across species, a phenomenon known as the "species effect"^{21–23}, which differs from cross-species comparisons at the organ/tissue level where transcriptomes often cluster primarily by organs/tissues rather than by species. A recent bioRxiv preprint²⁴ investigated this phenomenon and found that species effect observed when comparing single cell data across species can be largely explained by ortholog gene switching from specific to non-specific expression. As shown in **Fig. 3a**, our study also suggested that the cell-type expression profiles of orthologs tend to diverge across species as only ~ 3% of the 1-to-1 orthologs show all pairwise Pearson correlations greater than 0.9.

While cell types tend to cluster closely with other cell types from the same species rather than with homologous cell types from other species, we did notice that some cell types tend to be more conserved across species compared to others. **As shown in Response Fig. 3**, we quantified the euclidean distance between cell type transcriptomes across species in the PCA space and compared that for one cell type to the rest of the cell types as described by Andrijevic *et al.*²⁵ to determine significance. We observed large variations in cross-species distance of cell types. Compared to other cell types, vascular cell types (phloem and xylem) are more distinctly positioned from each other from different species, indicating that their transcriptomes are less conserved among species. On the contrary, the pericycle exhibits the lowest distance among species, suggesting the pericycle transcriptomes from different species are more conserved compared to other cell types.



Response Fig. 3. Cell type distance across four Brassicaceae species shown in Fig 5A. The distance was calculated for each pair of species under control condition. Asterisks indicate significance after Wilcoxon testing (n = 50) at adjusted p-value cutoff of 0.05

Minor comments:

L110: why is ABA concentration given as 'micromolar'

Response: We thank the reviewer for noticing this. We have changed this to “ μM ” in the revised manuscript at line 121 to keep it consistent with “mM” for NaCl treatment.

Figure 5D: I assume that the columns for each species correspond to the cell types in the previous panels. Please make this clear

Response: Yes, the columns in **Fig. 5d** represent the cell types and are colored based on the cell types. We have now clarified this in the legend.

L324: 'We observed 324 overall larger transcriptomic distinctiveness between species in response to 100 mM NaCl 325 compared to that of 5 μM ABA (Fig. 5E).' I am unsure how the authors came to this conclusion from the PCA figure. Please make this more clear.

Response: We thank the reviewer for this comment and apologize for the lack of clarity here. Each data point in **Fig. 5e** represents between-species transcriptomic distinctiveness for a given cell type under a given condition. To support our conclusion, we quantified the distance between treated and control samples in the PCA space shown in **Fig. 5e**. **As shown in the newly added Supplementary Fig. 26**, the distances between NaCl-treated (as indicated by the darker colors) and control samples are consistently greater than those between ABA-treated (as indicated by the lighter colors) and control samples across cell types. This suggested that NaCl treatment elicited a stronger impact on between-species transcriptomic distinctiveness compared to ABA treatment.

Reviewer #5 (Remarks to the Author):

In this study, Wang et al. generated comprehensive and valuable single-cell RNA-seq datasets from Brassicaceae roots, including stress-sensitive species, extremophytes, and a polyploid crop, under NaCl and ABA treatments. They also developed a framework for comparative analyses using scRNA-seq datasets. While the study provides insights into how stress responses diversify across cell types and species, several concerns remain:

1. Estimation of divergent responses: The authors rely heavily on previously published methods to infer gene expression divergence among species. However, it is unclear exactly how this was done and why specific thresholds (e.g., probability cutoffs) were chosen for further analyses. I also wonder whether the estimation approach, originally developed for mammalian cells, can be directly applied to crops.

Response: We thank the reviewer for raising this important point. As the reviewer noticed, in our study, expression divergence among 1-to-1 orthologs was inferred using a phylogenetically informed probabilistic framework adopted from Cardoso-Moreira *et al.*, (2019)²⁶. Briefly, we first clustered the orthologs based on their expression profiles across the ten defined root cell types using soft clustering (Mfuzz), which assigns probabilistic rather than binary membership to expression patterns. We then inferred, along the Brassicaceae phylogeny, the probability that an

ortholog changed its expression profile on a given branch, explicitly accounting for uncertainty in cluster assignment. The probability cutoff used to call divergently expressed orthologs (posterior probability ≥ 0.001) follows the original framework and was chosen to be intentionally stringent to minimize false positives driven by expression noise. We further validated the expression divergence of selected orthologs using another independent method HCR, which confirmed the divergent expression detected using this approach. Relaxing the probability cutoff increases the number of inferred changes but does not alter the qualitative conclusions regarding lineage-specific versus convergent divergence.

Although the framework was originally developed for mammalian systems, it is agnostic to organismal lineage as it relies only on 1) quantitative expression measurements across comparable biological contexts and 2) a resolved species phylogeny. In our case, both criteria are met: gene expression is measured across homologous root cell types, and the Brassicaceae phylogeny (especially the lineages to which species in our species belong) used has been well established. We therefore adapted the framework to our dataset by using the expression of orthologs across cell types as the evolutionary units and by restricting analyses to unambiguous 1-to-1 orthologs to avoid confounding effects from gene duplication. We have revised the Methods at lines 830-853 to more clearly describe the procedure, threshold choice, and the underlying assumptions.

2. Lack of quantitative measurements: Although the authors provide the numbers of orthogroups in most figures (e.g., Figures 1E and 3B), they proceed directly to discussing individual genes. This makes it difficult to assess the extent or proportion of genes affected or belonging to each group. For instance, in Figure 5F, the authors state that “cortex and atrichoblast transcriptomes significantly increased their similarity following NaCl treatment only between the two extremophytes.” However, without more quantitative data, this is not clearly supported by the current visualizations.

Response: We thank the reviewer for this important comment. For **Fig. 5f**, the reported increase in cortex and atrichoblast transcriptomic similarity between the two extremophyte was based on the comparison of transcriptomic distinctiveness between NaCl-treated and control conditions. More specifically, we observed that the distinctiveness of cortex and atrichoblast transcriptomes between *E. salsugineum* and *S. parvula*, but not between any other species pairs, was lower under NaCl treatment compared to the control. This indicated that the cortex and atrichoblast transcriptomes from *E. salsugineum* and *S. parvula* became more similar upon NaCl treatment. We have added a supplementary heatmap as **Supplementary Fig. 28** to explicitly show the quantitative shifts in transcriptomic dissimilarities for all species pairs and cell types, and pinpoint the specific comparisons highlighted in the manuscript.

More broadly, we have revised the Results section to include quantitative metrics whenever applicable. For example, we added the fraction of flavonoid biosynthetic genes at line 274 that were differentially expressed between cortex layers, the proportion of genes in the stress-

prepared clusters among all clustered orthologs at lines 461-471, and percentage of DEGs in each sub-genomes in *C. sativa* at lines 515-516.

3. Scope of ortholog comparisons: In most analyses, the authors focus on 1:1 orthologs (15,198), except in Figure 6, where at least 30,133 genes were included. This suggests that many claims are based on only one-third to one-half of the total genes in each species. It would be important to assess whether genes with one-to-many or many-to-many orthologous relationships show similar trends.

Response: We thank the reviewers for pointing this out. We've intentionally used the 15,198 1-to-1 orthologs for all the cross-species comparisons (including **Fig. 6 and its accompanying supplementary figures**) throughout our study to reduce the complexity of the study. For the comparisons shown in **Fig. 6**, transitions between DEGs and non-DEGs among the 1-to-1 orthologs under each treatment were first counted within each cell type and then summed across all cell types. If the transition between DEGs and non-DEGs for a given 1-to-1 ortholog was found in multiple cell types, we counted how many times it occurred across the cell types (not just once). Therefore, 6,848 and 30,133 transitions represent the total number of such events from all ten cell types defined in our study, and many of these transitions are shared among cell types. We have added numbers of transitions to **Supplementary Fig. 29** in the revised manuscript to clear the confusion.

We agree that it is important to determine if similar trends also apply to one-to-many or many-to-many orthologs. However, it becomes noisy to perform cross-species comparisons using these ortholog groups as it is difficult to identify the correct corresponding orthologs within these groups. To address the reviewer's comment under this restriction, we identified ortholog groups with more than one copy in only one species and tested if they exhibited similar evolutionary trends as the 1-to-1 orthologs. As shown in the newly added **Supplementary Fig. 30**, we observed nearly identical trends for these ortholog groups where lineage-specific gains of responses were consistently lower than expected, but lineage-specific losses of responses exceeded the expectation. This suggests that our finding about the evolutionary shifts between DEGs and non-DEGs also applies to duplicated ortholog groups. We have added these results to the revised manuscript at lines 413-416 (the corresponding method is added at lines 909-910) to provide additional support for our findings.

4. Interpretation of species and treatment effects: In the abstract, the authors claim that "cell-type contributions to stress responses differed both between species and across treatments." However, based on Figures 5 and 6, this conclusion appears largely descriptive and not strongly supported by comparative analyses. Specifically, the NaCl and ABA datasets were analyzed separately rather than jointly. Given the well-known cross-talk and overlapping responses between salt and ABA signaling, it would be valuable to identify genes showing shared or distinct responses to these two stresses across species and cell types. For example, a co-expression clustering approach combining both stress conditions could strengthen this point.

Response: We thank the reviewer for raising this constructive suggestion. We agree that disentangling the cross-talk from treatment-specific responses is critical for our argument. Our claim about cell type contributions to stress responses differed between species and treatments was primarily based on **Fig. 5b**, which demonstrated that the rank order of cell-type responsiveness significantly shifted between treatments in each species (e.g. while ABA triggered greatest transcriptomic remodeling in procambium in Arabidopsis, the effect of NaCl primarily concentrated in pericycle and cortex).

To further validate our claim, we performed a direct comparison of stress responses by identifying genes that showed shared or distinct responses to ABA and NaCl in all cell types across the species (**Supplementary Fig. 25**). This quantitative assessment revealed that genes shared between both stresses typically accounted for less than 50% of responsive genes from either treatment, with a notable exception in Arabidopsis where up to 78% of the ABA-responsive genes were shared with the NaCl stress response. The shared genes consistently constituted a much larger proportion of the ABA-responsive set than of the NaCl-responsive set. These together confirm that while cross-talk exists, the majority of the transcriptomic remodeling is indeed stress-specific, and salt stress activates not only ABA-mediated responses but also a massive additional set of ABA-independent pathways. We have added these results to the revised manuscript at lines 330-339 for additional support for our finding.

5. Analysis of ABF and ABA responses: As noted by the authors, ABF directly binds to ABA-responsive elements, so the overrepresentation of ABA-related events compared with salt treatment is not surprising. The authors argue that this may reflect a more complex ABA-mediated regulation during salt stress (Figure 6C), but this remains speculative. Performing DAP-seq with salt-specific transcription factors could provide a meaningful counterpart analysis. Additionally, the statistical test described in the Methods (lines 810–811: “The significance of differences in the percentage of divergently regulated orthologs between the two defined groups were determined with t.test function in R”) may not be appropriate, as t-tests assume normal distribution.

Response: We thank the reviewer for this insightful comment. We agree that having a counterpart analysis with DAP-seq for salt-specific transcription factors would be highly informative. However, generating *de novo* DAP-seq profiles for additional transcriptional factors is beyond the scope of the current study. We focused on ABFs specifically because high-quality reference cistrome data is publicly available, allowing us to robustly test the relationship between the divergence in ABF binding and transcriptional responses without generating new protein-DNA binding datasets. We wish to include DAP-seq for salt-specific transcription factors in our follow-up studies.

Regarding the statistical test used for **Fig. 6c**, we agree that a t-test may not be appropriate for percentage data that may not follow a normal distribution. To address this and ensure robustness, we have re-analyzed the differences in the percentages of divergently responsive orthologs using

the non-parametric Wilcoxon rank-sum test. The Wilcoxon test produced results fully consistent with those obtained from the t-tests, and therefore did not change the statistical significance or the interpretation of our findings. We have updated **Fig. 6c** and **Supplementary Fig. 33** with the Wilcoxon tests, and also the Methods section to reflect this change in the statistical methodology.

6. Shiny app performance: Although the authors acknowledged long loading times, the current Shiny apps are still difficult to use. Despite multiple attempts with different internet connections and browsers, I repeatedly encountered timeout errors.

Response: We thank the reviewer for trying out our browser and for bringing up this issue. To mitigate this issue, we have reduced the size of the data set used to initiate the browser and extended the initiation timeout. In our experience using Google Chrome and Microsoft Edge, loading [our project page](#) now generally takes less than 5 seconds, while loading the data pages for each species and the integrated atlas takes less than 1-2 minutes. If it takes longer to load, we suggest that the users clear the cache and cookies of the browser and try again.

Reviewer #5 (Remarks on code availability):

It's pretty clear to me.

Response: Thank you and we appreciate the confirmation.

References used in the response:

1. Shahan, R. *et al.* A single-cell Arabidopsis root atlas reveals developmental trajectories in wild-type and cell identity mutants. *Dev. Cell* **57**, 543–560.e9 (2022).
2. Kagale, S. *et al.* The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nat. Commun.* **5**, 3706 (2014).
3. Chaudhary, R. *et al.* Sequencing of *Camelina neglecta*, a diploid progenitor of the hexaploid oilseed *Camelina sativa*. *Plant Biotechnol. J.* **21**, 521–535 (2023).
4. Li, H. *et al.* Effective root responses to salinity stress include maintained cell expansion and carbon allocation. *New Phytol.* **238**, 1942–1956 (2023).
5. Tran, K.-N. *et al.* Balancing growth amidst salt stress - lifestyle perspectives from the extremophyte model *Schrenkiella parvula*. *Plant J.* **116**, 921–941 (2023).
6. Zhu, M. *et al.* Single-cell transcriptomics reveal how root tissues adapt to soil stress. *Nature* (2025) doi:[10.1038/s41586-025-08941-z](https://doi.org/10.1038/s41586-025-08941-z).
7. Cervantes-Pérez, S. A. *et al.* Cell-specific pathways recruited for symbiotic nodulation in the *Medicago truncatula* legume. *Mol. Plant* **15**, 1868–1888 (2022).
8. Skinnider, M. A. *et al.* Cell type prioritization in single-cell data. *Nat. Biotechnol.* **39**, 30–34 (2021).

9. Squair, J. W. *et al.* Confronting false discoveries in single-cell differential expression. *Nat. Commun.* **12**, 5692 (2021).
10. Murphy, A. E. & Skene, N. G. A balanced measure shows superior performance of pseudobulk methods in single-cell RNA-sequencing analysis. *Nat. Commun.* **13**, 7851 (2022).
11. Guillotin, B. *et al.* A pan-grass transcriptome reveals patterns of cellular divergence in crops. *Nature* **617**, 785–791 (2023).
12. Agati, G. *et al.* Antioxidants by nature: an ancient feature at the heart of flavonoids' multifunctionality. *New Phytol.* **245**, 11–26 (2025).
13. Daryanavard, H., Postiglione, A. E., Mühlemann, J. K. & Muday, G. K. Flavonols modulate plant development, signaling, and stress responses. *Curr. Opin. Plant Biol.* **72**, 102350 (2023).
14. Brunetti, C., Fini, A., Sebastiani, F., Gori, A. & Tattini, M. Modulation of phytohormone signaling: A primary function of flavonoids in plant-environment interactions. *Front. Plant Sci.* **9**, 1042 (2018).
15. Mao, Y., Luo, J. & Cai, Z. Biosynthesis and regulatory mechanisms of plant flavonoids: A review. *Plants* **14**, 1847 (2025).
16. Buer, C. S. & Muday, G. K. The transparent testa4 mutation prevents flavonoid synthesis and alters auxin transport and the response of Arabidopsis roots to gravity and light. *Plant Cell* **16**, 1191–1205 (2004).
17. Miyashima, S., Sebastian, J., Lee, J.-Y. & Helariutta, Y. Stem cell function during plant vascular development. *EMBO J.* **32**, 178–193 (2013).
18. Lucas, W. J. *et al.* The plant vascular system: evolution, development and functions. *J. Integr. Plant Biol.* **55**, 294–388 (2013).
19. Agustí, J. & Blázquez, M. A. Plant vascular development: mechanisms and environmental regulation. *Cell. Mol. Life Sci.* **77**, 3711–3728 (2020).
20. Zeng, M. *et al.* Stem cell lineage in body layer specialization and vascular patterning of rice root and leaf. *Sci. Bull. (Beijing)* **61**, 847–858 (2016).
21. Shafer, M. E. R. Cross-species analysis of single-cell transcriptomic data. *Front. Cell Dev. Biol.* **7**, 175 (2019).
22. Luecken, M. D. *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022).
23. Song, Y., Miao, Z., Brazma, A. & Papatheodorou, I. Benchmarking strategies for cross-species integration of single-cell RNA sequencing data. *Nat. Commun.* **14**, 6495 (2023).
24. Song, Y., Arendt, D., Papatheodorou, I. & Brazma, A. Revising the ortholog conjecture in cross-species comparison of scRNA-seq data. *bioRxiv* (2024) doi:[10.1101/2024.06.21.600109](https://doi.org/10.1101/2024.06.21.600109).
25. Andrijevic, D. *et al.* Cellular recovery after prolonged warm ischaemia of the whole body. *Nature* **608**, 405–412 (2022).

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