

Chromosome-scale genomes and population resequencing resolve subgenome diversity and halophyte adaptation in *Salicornia*

Corresponding Author: Dr Vanessa Melino

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)

Wang et al. provide valuable genomic resources, including six high-quality *Salicornia* genomes and a large resequencing panel. However, the manuscript falls short of the analytical depth and mechanistic insight expected for a Nature Communications research article. Although I was excited to see such comprehensive data, much of the text remains descriptive, mainly focused on assemblies, basic comparisons, and resource generation, without proposing new hypotheses or successfully uncovering biological mechanisms. The authors must substantially deepen their analyses and interpretations to transform this work from a data announcement into a study that delivers genuine biological discovery.

Major concerns:

1. The study does not fully exploit its genomic data to explore gene families that could underlie halophyte adaptations. A comprehensive analysis of gene family expansions or contractions (especially those related to salinity tolerance) is missing. For instance, ion transporters, osmolyte biosynthesis enzymes, stress-signaling proteins, and transcription factor families known to be important for salt stress are not analyzed in detail. It would greatly strengthen the manuscript to identify if *Salicornia* species have unique expansions of such gene families or positively selected variants in these genes. Currently, the authors only mention clustering genes with OrthoFinder for a species tree but do not discuss any functional gene family results. Targeted analysis (e.g., how many NHX-type Na^+/H^+ antiporters or HKT transporters exist in *Salicornia* vs. related glycophytes, or whether halophyte-specific genes like LEA proteins or ion channels are expanded) would provide biological insight. Without this, the genomic resource is not leveraged to explain *Salicornia*'s salt tolerance traits.
2. Given six high-quality *Salicornia* genomes, the lack of a thorough structural variation analysis is a major gap. The authors mention identifying three chromosomal translocations in the tetraploid *S. bigelovii* relative to diploids (validated by FISH), with no gene losses at the breakpoints, but there is no systematic genome-wide SV survey. The manuscript should examine structural variants such as large insertions/deletions, inversions, copy-number variations, or presence/absence of genes across the species. Such analysis could uncover structural changes that contribute to species-specific adaptations (for example, gene duplications or deletions related to stress response). Currently, the focus on just a few translocations means potential SV impacts on salt tolerance or other traits remain unexplored. A comprehensive SV analysis across the genomes is needed to fully capitalize on the data.
3. The population genomic analysis should be extended beyond basic clustering to identify genetic variation associated with salt tolerance or other adaptive traits. The authors present phylogenetic and structure analyses for the resequencing panel (Fig. 4 and Fig. 5), showing taxonomic groupings and some admixture. However, there is no attempt to connect these patterns to phenotypic or environmental variables. For a study centered on salt tolerance, it is crucial to ask: do certain populations or clades harbor alleles linked to higher salt tolerance? The manuscript should include analyses that probe adaptive divergence – for example, genome scans for differentiation between populations from high-salinity vs. lower-salinity environments, or association tests linking genomic variants to any available trait data. As is, the population analysis is purely descriptive (who is related to whom) and does not reveal any why or how regarding adaptation.
4. The authors report population structure and admixture patterns in diploid *Salicornia* (e.g., mixed ancestry between *S. europaea* (Israel/Egypt) and *S. persica* ssp. *iranica* in Fig. 5a). However, they do not formally assess gene flow between species or populations. Given the misclassified accessions and close relationships, tests for introgression would be highly informative. The manuscript should include analysis of gene flow – for instance, D-statistics (ABBA-BABA tests) or other introgression methods – to determine if hybridization has occurred between certain lineages. This would add evolutionary

insight beyond the STRUCTURE analysis. Right now, the text infers admixture from population structure ($K=2-4$). Including such analyses would strengthen the conclusions about inter-species relationships and the evolutionary history of *Salicornia*.

5. A major omission is the lack of analysis for selective sweeps or loci under natural selection related to salt stress. The study provides a wealth of genomic variation data, but the authors have not mined it for signals of adaptation. To merit publication in a high-impact venue, the authors should perform genome-wide selection scans (e.g., F_{ST} outlier analyses, XP-CLR, XP-EHH, or $\Delta\pi$ scans between groups). For example, if coastal vs. inland populations were sampled, are there regions of the genome with high differentiation that might indicate local adaptation to salinity? The text briefly notes F_{ST} values (e.g., $F_{ST} \approx 0.59$ between *S. europaea* (Israel) and *S. persica iranica*), but it does not go further to identify which genes or regions contribute to such divergence. The authors should identify any candidate genomic regions under selection and, importantly, highlight if those regions contain known abiotic stress response genes. As an illustration, if genes like *HKT1;1* or *NHX1* (known Na^+ transporters) or osmoprotectant synthesis genes show signatures of a selective sweep in high-salinity populations, that would directly link the genomic data to salt tolerance. Without such analysis, the paper misses an opportunity to connect genotype to adaptive phenotype.

6. The RNA-seq differential expression analysis is not convincing in its current form. It appears the authors relied on a public transcriptome dataset rather than conducting a new, controlled experiment tailored to their hypotheses. As a result, the differential expression results are of limited scope and relevance. For instance, focusing on *Salicornia bigelovii* alone and using data not specifically designed to probe salt stress mechanisms makes it hard to draw firm conclusions. The manuscript states it identified genes associated with high shoot Na^+ accumulation via “comparative transcriptomic analyses” (abstract) but this was done by extracting DEGs unique to *S. bigelovii* from an existing dataset. This approach is suboptimal: factors other than salt tolerance could differ between *S. bigelovii* and others, and without a controlled experiment (same conditions, replicates, time points), the results are merely correlative. To strengthen this aspect, the authors should perform their own salt stress experiment or at least validate the key findings with additional data. The decision to highlight only two genes – *SOS1* and *SOS2* – in the transcriptomic analysis is not well-justified. These genes (components of the SOS salt tolerance pathway) are indeed important, but the exclusive focus on them suggests a very narrow analysis. The authors report that *S. bigelovii* uniquely upregulates *SOS1* and has high baseline expression of *SOS1* and *SOS2* compared to quinoa or spinach. However, why limit the discussion to these two genes? There are many other salt-responsive genes (ion transporters, osmolyte synthases, LEA proteins, ROS-scavenging enzymes, transcription factors like *DREB*, *bZIP*, etc.) that could have been examined. If the intent was to validate the RNA-seq findings, focusing on only *SOS1/2* gives an incomplete picture. The manuscript would be far more compelling if a broader set of candidate genes from the DEGs were analyzed or at least if the authors explained that *SOS1/2* were chosen a priori due to known importance. In addition, some validation of their expression (for example, qPCR in multiple species or conditions) would help. As is, the emphasis on just two genes appears arbitrary and insufficient, leaving the reader wondering if other crucial genes were overlooked.

Minor concerns:

1. Which data did the authors use to generate the K_s value in Figure 3b? And it would be better to use the model fit to detect the peak.
2. Line 265-266, I cannot find the evidence support this conclusion.
3. Line 309, is it “bottleneck” or “founder effect”?
4. How many single-copy gene groups have been identified in this study? Did the MCMCtree’s result have been tested by Tracer?
5. The authors write in the results that they used the IAGS to rebuild the ancestor genomes, but they did not write the detail methods in the “Methods” part. What about the evaluation results for this ancestor genome?
6. the results in Extended Data Fig. 10, if the authors really want to explore something about the distribution pattern of SNP, they may separate the variations sets into different groups according to the Figure 5a. The authors indeed calculated the π and F_{st} , which only provided very limited information, and where is the selection sweep analysis? I do believe that the authors could explain more based on their data.
7. In line 79-81, the authors said that there were several questions such as “evolution, gene flow and gene regulatory mechanisms”. But through this study, these questions still remain. Then what is the scientific questions and the main purpose for this paper?
8. Line 87. What did the author mean for the “diversity”? What does the authors mean for “diversity assembled”?
9. Line 113-115. I don’t think this part belongs to the “Results”.
10. The authors propose a mechanism wherein *Salicornia* (especially *S. bigelovii*) achieves extreme salt tolerance via constitutively high expression of Na^+ transport genes like *SOS1*, enabling continuous sequestration of Na^+ into vacuoles. While this is a plausible hypothesis – and is consistent with prior studies in *Salicornia*. the manuscript’s own data do not definitively prove this mechanism. There is no functional experiment (e.g., transgenic or mutant analysis) demonstrating that higher *SOS1* or *SOS2* expression causally leads to greater salt tolerance. Thus, any claims about “mechanisms” should be toned down.
11. The manuscript asserts that these genomic resources “lay the foundation for the domestication of *Salicornia* as a climate-resilient crop”. While the data (genome assemblies and a diversity panel) are undoubtedly valuable for crop development, this claim goes beyond what is demonstrated. The paper does not identify or improve any domestication traits (e.g., it does not discuss yield, seed shattering, growth rate, etc., which are critical for domestication). Claiming to pave the way for domestication is premature without showing how the genomic insights will be applied to breeding.

(Remarks on code availability)

The computational analysis scripts are currently incomplete, which hinders reproducibility and review. Please provide the full, end-to-end code for each analysis to enable a thorough evaluation.

Reviewer #2

(Remarks to the Author)

The authors submitted a manuscript with the title "From Wild Halophyte to Future Crop: Genomic Insights into *Salicornia*". The study examines the genomes of several *Salicornia* species. As part of the *Amaranthaceae* family, *Salicornia* is appreciated for its edible succulent stems and oil-rich seeds, though enhancement of its genetics has been challenged by limited genomic data. Chromosome-scale, annotated genome assemblies for six *Salicornia* species have been provided, with four distinct subgenomes identified within the genus. Genes linked to high shoot Na⁺ accumulation, a crucial mechanism for salt tolerance, have been discovered through comparative transcriptomic analyses. Interspecies relationships and misclassification at both accession and species levels are illuminated by phylogenetic clustering and population structure analyses using a globally representative resequencing panel. With these insights, a solid basis is formed for *Salicornia*'s development as a climate-resilient crop tailored for saltwater agriculture.

The sequencing of six *Salicornia* genomes is indeed excellent news for the scientific community, offering a wealth of data that will hopefully soon be publicly accessible. However, a primary concern is that, while the manuscript presents a significant amount of data, the information is not thoroughly interconnected, evaluated, or deeply interpreted. This omission results in several major and minor contradictions throughout the manuscript. It's important to recognize that journals like *Nature Communications* have space constraints that limit the extent to which results can be elaborated in detail.

Nevertheless, the authors have a crucial responsibility to strategically focus on and highlight the most significant findings while also addressing and not overlooking any contradictory information. By improving the integration, evaluation, and interpretation of the collected data, the manuscript can effectively convey its contributions to the field and maintain a coherent narrative. This approach will ensure that the results are presented with clarity and precision, thus enhancing the overall impact of the study.

Note: Sequence of authors in the invitation email/submission email is not the same as in the paper itself with respect to the author Vanessa Melino.

Abstract and title

The current title of the manuscript is somewhat cryptic and could be more descriptive. To enhance clarity, consider revising it to better reflect the focus on genome data and its implications for developing *Salicornia* spp. as a crop plant. While the manuscript provides valuable insights into genome data, which is certainly of high interest for advancing *Salicornia* as a crop, it lacks concrete facts on its development as a crop plant. To address this gap, consider integrating more detailed information pertinent to crop development. Additionally, the abstract could be strengthened by incorporating more quantitative data. At present, it is quite general, but could be elevated by including key numbers and precise information. Look to abstracts from other sequencing studies as examples, which often highlight quantitative findings to convey the significance of their research upfront. By implementing these revisions, the manuscript would not only become more accessible and informative but also underscore its relevance to the advancement of *Salicornia* spp. as a crop species.

Introduction

The main part provides a good introduction into the topic. Some general statements are taken unquestioned from other studies, some references are certainly missing (see later in this review). For this type of journal, *Nature Communications*, the formulation of biological hypotheses is necessary, at least from my point of view. Otherwise, it remains a descriptive sequencing paper that could be published in more suitable journals. Although the approach and the data produced are certainly of high interest for future studies.

Materials and Methods

Several preprints are cited that have not yet been peer-reviewed, they might never be fully published. Some details are missing (e.g., conditions in the greenhouse). The homogeneity of seed material is not mentioned. Therefore, it is important to mention the number of plants that were used/pooled for the analyses, for example for the RNA extraction.

In the case of the SNP data, large quantities were apparently collected, which later could scarcely be explored in detail.

As I am not an expert in bioinformatics, I am not commenting on the methods evaluating the sequencing data.

Results

The authors collected a lot of interesting and relevant data. Some results as the ploidy levels might be so important for this particular study to present them in the main part. For example, in the appendix it is shown that the 2C genome size of *Salicornia europaea* (France) does not correspond to the known genome sizes of *S. europaea*. This difference is also visible in the phylogenetic tree. Therefore, the sequenced genome should better not be labeled as *S. europaea* in contradiction to the literature mentioned earlier. It could indicate to another species and a few clear traits exist for species identification, aside from the ETS marker sequences (Kadereit et al. 2012) and genome size (Zonneveld 2019). In lines 227-229, it states: "*S. procumbens* accessions formed a distinct clade within the larger tetraploid *S. europaea* clade from France (Fig. 4 and Extended Data Fig. 7)." One could ask why the species from France was not classified as *S. procumbens*?

What was the reasoning to select these species for the phylogenetic tree, please add some information.

Line 93-95 explicitly criticizes that many species have been labeled as *S. europaea* at various locations. Line 122 mentions the supposedly tetraploid *S. europaea* species, which according to Kadereit et al. 2012 cannot exist, as only the genetically similar *S. perennans* species include tetraploid types; moreover, *S. procumbens* is another known European tetraploid species. The situation is unclear because Table 1 indicates a 2C genome size of 1129 Mb (which corresponds to *S. perennans* in Zonneveld 2019), but the measured pg content isn't mentioned, which should have been 2.309 pg correspondingly, suggesting the tetraploid *S. procumbens* species. The fact that speaks against *S. procumbens*, as

mentioned earlier, is that *S. europaea* (France) contains somewhat similar subgenomes compared to *S. europaea* (and other diploid species), indicating an assignment to the diploid *S. perennans* species, though the latter isn't mentioned either. While I somewhat understand the line of reasoning, I can't draw clear conclusions besides noting these contradictions due to the missing pg content.

When referring to *Salicornia* as an excellent protein source, it is surely the seeds being referred to, which, to my knowledge, should not be consumed in large quantities due to the saponins they contain. (Line 66).

There is no clear distinction between *Salicornia*, *Sarcocornia*, and *Tectocornia*. It should be mentioned that these are different genera (Line 68).

Lines 273-276: This should be part of the discussion.

"Smaller collections of *S. europaea* from China and the Netherlands also formed a separate clade, clustering with *S. brachiata* (Fig. 4)." So, are Dutch, Indian, and Chinese species closely related? I find no conclusions from this result.

Line 412: "Cross-species hybridization has historically been difficult." References?

How do sequenced genomes help overcome cross-breeding incompatibility? If anything, more data has been gathered on why certain species are not crossable (Line 414).

Lines 417-429: All the effort of intra-family transcriptome comparison just to bring up an already known pathway (SOS) and specifically the SOS1 transporter, which has already been frequently discussed in connection with *Salicornia*? The potential eco-physiological adaptations (larger vacuoles; habitat specialization) due to ploidy levels (Kaligarić et al. 2008) were not discussed, even though collecting wild types provided an opportunity for closer examination: "We could add that annual tetraploids, derived from diploids, then occupied the even more extreme, previously empty space of mudflats along shallow coasts, where tidal activity is a limiting growing factor, with its periodicity of flooding." Already Noble et al. (1992) report that tetraploids tend to be more characteristic for the lower parts of the marshes." Why did you select *S. biglovii* for more detailed gene expression analysis? A comparison of all six species would have been interesting. The authors could think about hybridization events to explain some of the data.

There are some more examples like these. Therefore, a careful revision of each sentence to draw final (or prefinal) conclusion based on a combination of the various data sets is highly recommended.

Discussion

A lot of data have been collected, but only few are discussed. The first paragraph of the discussion can be deleted as everything was written in the introduction already. That gives some space for a deeper discussion. The biological results are only very shortly discussed. Many relevant comparable facts have been already published and that means several references are missing.

What is the exact meaning of subgenomes? Can you draw your conclusions based on the six genomes? I assume that more data are needed to prove your hypothesis.

It is highly recommended to discuss your data that means to compare to other genera or families.

Lines 394 – 398: Actually, this is known already and could be deduced from previous data.

Lines 418-420: As you are describing gene expression, please write gene names in Italics.

Line 427: Salt Overly Sensitive SOS pathway: The abbreviation is used already several times and now it is introduced, please correct.

Minor

There are many writing and formatting errors that need to be very carefully corrected, too many to point them out individually. Also the list of references contains many writing and formatting and fact (wrong years) errors.

The Bioproject data were not yet available but these will be hopefully available with the acceptance of the manuscript.

I personally, do not recommend to write in the first person but in a passive way.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

The paper mainly reports six genomes of *Salicornia* species and a basic population genetics analysis that partially resolves relationships within the genus. Despite the ambitious title "From Wild Halophyte to Future Crop: Genomic Insights into *Salicornia*", the paper does not demonstrate how these resources will accelerate *Salicornia*'s domestication or improve agronomic traits. Moreover, the genomic story lacks depth and novelty, leaving significant opportunities to strengthen both its scientific impact and its relevance to genome evolution.

More comments are listed below.

Line 83-87. The authors criticize earlier *Salicornia* genomic studies for their low assembly quality and incomplete

annotations. The authors should provide concrete evidence to support this claim. Additionally, there is a misreferenced citation.

The study lacks paired-end short-read sequencing, which is essential for genome polishing and consequently fails to perform the necessary polishing to correct residual errors.

Given the multiple ploidy levels among the sampled species, paired-end short-read data are also indispensable for accurate genome size estimation and ploidy inference, thereby ensuring a more complete and robust body of evidence.
[<https://www.nature.com/articles/s41467-020-14998-3>]

The karyotype presentation in Extended Data Fig. 3 is unclear and lacks sufficient resolution.

Line 139-141. The authors note that karyotyping of *S. europaea* (France) was unsuccessful. Please consider applying multiple techniques and repeating the assays to achieve a successful karyotype.

Line 166-170. For subgenome homology analysis, it is standard practice to include appropriate outgroup taxa and to build phylogenetic trees using aligned CDS data, rather than relying solely on OrthoFinder's fast-tree mode.

In Fig. 3b, why do intra-species paralogous genes in the diploid taxa exhibit a recent Ks peak, particularly in *S. brachiata*?

In Fig. 3a, although Line 205-206 describes "the divergence of the diploid progenitors", the figure provides no annotation to clarify this, which may lead readers to misinterpret the two circles as marking the timing of polyploidization events. Moreover, directly plotting progenitor divergence times on the phylogenetic tree is unconventional and should be reconsidered.

Line 208-212. The results in Fig. 2c and Fig. 3c are inconsistent: Fig. 2c reveals a larger number of chromosomal structural rearrangements, whereas they should display the same overall pattern.

The population genetics analysis is relatively simple and conventional, maybe constrained by low sequencing depth. They fail to resolve the Salicornioideae subfamily taxonomy or to clarify the origin and evolutionary history of the genus.

Line 323. The gene function analysis lacks innovation, relying solely on basic transcriptome profiling and the selection of a handful of differentially expressed genes. By contrast, the earlier Nature Communications paper, titled "SOS1 tonoplast neo-localization and the RGG protein SALTY are important in the extreme salinity tolerance of *Salicornia bigelovii*" offers far deeper mechanistic insights and experimental logic.

Line 337-344 FPKM is a relative measure of gene expression, and its application may lead to inaccurate results when used to compare expression levels across different species.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The revised manuscript is further improved relative to the first submission, moving the work beyond a "resource announcement" toward a more hypothesis-driven evolutionary genomics study. However, several issues still limit interpretability and should be addressed before the manuscript can be accepted.

Major concerns

1. Selection inference at the OSCA locus requires additional support.

The OSCA region is presented as a central adaptive result, with contrasting Tajima's D between *Salicornia* sp. France and *S. procumbens* interpreted as evidence for a selective sweep versus balancing selection. However, Tajima's D is sensitive to demography, population structure, uneven sampling, and genotyping/mapping bias, and on its own is insufficient to conclude balancing selection, particularly in complex polyploid-derived systems. Please provide complementary evidence in the OSCA region, such as local nucleotide diversity (π), a sliding-window F_{ST} scan between the two tetraploid groups, or a composite-likelihood (CLR) type statistic. In addition, please present allele-frequency patterns and local haplotype structure around OSCA to demonstrate that the signal is not driven by local mapping bias or missing data.

2. SNP calling strategy for the tetraploid A-subgenome should be described more transparently.

The manuscript states that SNP-based analyses were conducted on the A-subgenome of the tetraploids to compute π and Tajima's D, but the calling procedures and filtering criteria are not sufficiently detailed in the main text. Please clarify: (i) which reference was used (i.e., the source assembly and specific A-subgenome chromosome set), (ii) how homeologous mapping was prevented or minimized (e.g., unique-mapping thresholds, masking strategy, mapping-quality filters), (iii)

whether genotypes were treated as diploid when analyzing the A-subgenome and the justification for this assumption, (iv) the full filtering criteria applied to derive the tetraploid A-subgenome SNP set (e.g., QUAL/DP/MQ thresholds, missingness, MAF), and (v) the sample size (n) used for π and Tajima's D calculations (ideally reported in the relevant figure legends).

Minor concerns

1. Please clarify and justify the outgroup choice used in each set of tests and ensure consistency across analyses.
2. Please soften deterministic phrasing such as "balancing selection" and "rapid fixation of an optimal OSCA allele," as these are strong conclusions without demography-aware testing.
3. The revised manuscript still contains minor typographical issues (e.g., duplicated words such as "Notably, Notably," spelling inconsistencies like "S. europea" vs "S. europaea," and minor spacing/formatting problems). A careful final proofread is recommended.
4. Please also provide the CDS and protein fasta files of each genome assembly in the DRYAD.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

The authors have made revisions, and I appreciate the effort they have put into addressing the reviewers' comments. However, some of the changes appear overly expedient and, in my view, have even compromised the overall quality and integrity of the manuscript. In particular, several reviewers raised clear concerns regarding the RNA-seq analysis of the SOS1/2 genes, noting significant methodological issues. Instead of properly addressing these concerns, the authors have chosen to remove the relevant results entirely. This approach seems overly simplistic and somewhat evasive. Moreover, as the authors emphasize in both the title and throughout the manuscript, the halophytic characteristics of the genus *Salicornia*, the removal of this key functional analysis, substantially weakens the central biological narrative. Replacing it with a relatively simple comparison of gene family copy numbers is insufficient to support the claims made, particularly for a study aiming to meet the standards of Nature Communications.

More comments are listed below.

For the subgenome homology analysis, although the authors state that they used CDS data, the CDS sequences appear to have been directly aligned and used for phylogenetic inference. This approach is not sufficiently rigorous. Direct alignment of CDS can introduce alignment errors due to frame shifts and codon structure. A more reliable strategy is to first align the corresponding protein sequences of single-copy orthologs, and then translate the protein alignment to CDS alignments using PAL2NAL. After that, ambiguously aligned regions should be filtered out before tree construction. This workflow ensures codon-aware alignment quality and reduces phylogenetic artifacts. All phylogenetic analyses in the manuscript should follow this more rigorous strategy to ensure robustness and reproducibility of the inferred evolutionary relationships.

The authors attribute the recent Ks peak of intraspecies paralogous genes in the diploid taxa, particularly in *S. brachiata*, to tandem duplication events and provide Supplementary Fig. 4 as supporting evidence. However, this explanation is not clearly described or discussed in the main manuscript. Given that the presence of a recent Ks peak in diploid taxa could easily lead readers to speculate about additional whole-genome duplication events or other large-scale processes, it is important that the authors explicitly clarify this point. I recommend that the explanation regarding tandem duplication be incorporated into the relevant Results section or discussed more thoroughly in the Discussion to avoid confusion.

In Supplementary Note 1, the dot plots shown in Figures 1 and 2 display a lot of background noise. Please check it and use a reliable cut-off.

There appears to be an inconsistency in the use of italics for taxonomic names in both the text and figures. In several places, *Salicornia* and some Latin species names are not italicized (e.g., the second paragraph of Supplementary Note 1). Please carefully read the manuscript and all supplementary materials to ensure consistent formatting of genus and species names. In addition, for names such as *S. persica* ssp. *iranica*, please note that "ssp." should not be italicized, whereas the genus/species epithet and the subspecies epithet should be italicized.

Comments on the authors' response to Reviewer #2

The authors have made a relatively serious effort in responding to Reviewer #2. However, there remain two points that require further attention and clarification.

(1) On the response to the reviewer's comments regarding the Introduction

In the Response, the authors emphasize the question of whether genomic changes such as structural variation, copy-number variation, and gene-family remodeling are associated with adaptation to extreme saline environments. In the revised manuscript, this is reflected in the aim: "(iv) to identify candidate genes associated with salinity-related traits through comparative genomics within *Salicornia* and across the *Amaranthaceae*".

However, in practice, the results appear rather limited. The selection analyses seem to identify only a small number of candidates (e.g., *OSCA* gene family), followed by a brief discussion linking these to soil electrical conductivity (which presumably reflects salinity). As a result, this part of the manuscript has become somewhat underdeveloped and lacks depth.

(2) On the response to the reviewer's comments regarding transcriptome analysis and species selection

Reviewer #2 raised several important points regarding the intra-family transcriptome comparison ("Lines 417-429: All the effort of intra-family transcriptome comparison ..."), including the question "Why did you select *S. biglovii* for more detailed gene expression analysis?". In the current Response, the authors emphasize practical limitations. However, the key question regarding the rationale for selecting *S. biglovii* for detailed expression analysis has not been directly addressed.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have made a serious effort to address my previous concern about the *OSCA* locus, and the revised analyses are useful. The down-sampling control helps reduce concern about unequal sample sizes, the local nucleotide diversity profiles provide an additional layer of evidence beyond Tajima's *D*, and the remapping to an independent *S. ramosissima* reference is a helpful attempt to rule out reference-specific artefacts. The local haplotype-block visualization is also informative and strengthens the case that the *OSCA* region shows an unusual population-genetic pattern.

However, I do not think this issue has been fully resolved. My original concern was not only that Tajima's *D* required support from additional summary statistics, but also that the interpretation of balancing selection should be treated cautiously in a polyploid-derived system where demography, population structure, local structural variation, missingness, and residual homeologous mapping can all influence allele-frequency-based statistics.

The new analyses establish *OSCA* as a more credible candidate region, but they do not yet demonstrate balancing selection. In particular, I do not see the requested sliding-window *F_{ST}* analysis or a CLR-type selection scan across the *OSCA* region. These analyses would be important because the manuscript uses this locus as a central adaptive example. Without an independent statistic for local differentiation or selection, the current evidence is better described as a contrasting population-genetic signal or a selection-like signature, rather than as direct evidence that balancing selection is actively maintaining *OSCA* haplotypes.

I therefore recommend that the authors either add further quantitative support, such as a local *F_{ST}* scan and ideally a CLR/XP-CLR/RAiSD/SweeD-type statistic, or substantially soften the interpretation. A more appropriate conclusion would be that the *OSCA* region shows signatures consistent with divergent selection, or that it represents a candidate locus requiring further validation, rather than presenting it as a demonstrated case of balancing selection.

The clarification of the A-subgenome SNP-calling strategy is also an improvement, but several technical details should still be made more explicit. The authors should report the mapping-quality threshold, per-sample depth filters, missingness and heterozygosity summaries specifically for the *OSCA* region, and the sample sizes used for each π and Tajima's *D* calculation in the relevant figure legends. I would also encourage the authors to explain why HISAT2 was used for WGS read mapping, as this is not the most conventional choice for genomic resequencing data. For example, are there any studies or references to support that HISAT2 is the better choice in this case?

(Remarks on code availability)

The codes used for the revised and newly added analyses appear to be missing. Given that several key conclusions now rely on additional analyses, including down-sampling, nucleotide diversity scans, remapping to an independent reference, and local haplotype visualization, the corresponding scripts and parameter settings should be made available.

In addition, the code currently provided on GitHub would benefit from clearer documentation and annotation. At present, it is difficult for other researchers to understand the required inputs, software versions, key parameters, expected outputs, and the order in which the scripts should be run. I recommend that the authors provide a more complete and better annotated workflow, including example commands or a README file, so that the revised analyses can be reproduced by readers.

Reviewer #3

(Remarks to the Author)

The authors have revised the manuscript and adequately addressed my concerns. I have no further comments.

(Remarks on code availability)

Version 3:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Thanks for the efforts. Happy to see this manuscript publish online soon.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate 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/>

Point-by-point responses are provided below. We have supplied both (i) a manuscript with tracked changes and (ii) a clean revised manuscript with all changes accepted. Please note that a number of additional refinements are present only in the clean version, including relocation of some main text to the Supplementary Notes and minor grammatical or stylistic edits. These final adjustments were made to improve clarity and structure following substantial revisions and were not fully captured in the tracked-changes file due to the complexity of managing extensive edits in that format. We recommend reviewing the Main Manuscript Clean Version.pdf.

Reviewer #1 (Remarks to the Author):

Wang et al. provide valuable genomic resources, including six high-quality *Salicornia* genomes and a large resequencing panel. However, the manuscript falls short of the analytical depth and mechanistic insight expected for a Nature Communications research article. Although I was excited to see such comprehensive data, much of the text remains descriptive, mainly focused on assemblies, basic comparisons, and resource generation, without proposing new hypotheses or successfully uncovering biological mechanisms. The authors must substantially deepen their analyses and interpretations to transform this work from a data announcement into a study that delivers genuine biological discovery.

Major concerns:

1. The study does not fully exploit its genomic data to explore gene families that could underlie halophyte adaptations. A comprehensive analysis of gene family expansions or contractions (especially those related to salinity tolerance) is missing. For instance, ion transporters, osmolyte biosynthesis enzymes, stress-signaling proteins, and transcription factor families known to be important for salt stress are not analyzed in detail. It would greatly strengthen the manuscript to identify if *Salicornia* species have unique expansions of such gene families or positively selected variants in these genes. Currently, the authors only mention clustering genes with OrthoFinder for a species tree but do not discuss any functional gene family results. Targeted analysis (e.g., how many NHX-type Na⁺/H⁺ antiporters or HKT transporters exist in *Salicornia* vs. related glycophytes, or whether halophyte-specific genes like LEA proteins or ion channels are expanded) would provide biological insight. Without this, the genomic resource is not leveraged to explain *Salicornia*'s salt tolerance traits.

Response: We thank the reviewer for this constructive suggestion. In response, we substantially expanded our comparative genomic analyses to systematically examine gene family evolution associated with salinity tolerance and halophytic adaptation. The relevant updates to the methods and results of the analyses described below are now in the Materials and Methods under the sub-heading “*Gene family identification and phylogenetic analysis*”, and Results under the sub-heading “*Comparative genomic analyses reveal extensive structural variation and gene family turnover in *Salicornia**” and Discussion under the sub-heading “*Comparative genomics of salinity-related gene families*”.

First, we performed a targeted comparative analysis of key ion transport and salt-stress-associated gene families that have been extensively implicated in salinity responses in plants. These included

Na⁺/H⁺ exchangers (NHX; three classes), high-affinity K⁺ transporters (HKT), calcineurin B-like proteins (CBL), chloride channels (CLC) and slow anion channel-associated proteins (SLAC/SLAH). Gene family copy numbers were identified across six *Salicornia* species and 11 additional Amaranthaceae species, with *Arabidopsis thaliana* included as an outgroup, using a combined BLAST-based homology search and hidden Markov model (HMM) profile approach.

This analysis revealed that none of these canonical ion transport or stress-signalling families exhibited marked copy-number expansion within the *Salicornia* genus relative to other Amaranthaceae species. These results indicate that extreme salt tolerance in *Salicornia* is unlikely to be driven by large-scale expansion of classical transporter families, but instead may rely on alternative evolutionary mechanisms. The relevant results are presented in **Supplementary Figs. 8–12** and copy-number statistics provided in **Supplementary Tables 13–15**.

Second, following this reviewer's comment and the suggestion of Reviewer 3, we performed a genome-wide gene family expansion and contraction analysis using CAFE on a subgenome-resolved phylogeny. This analysis identified 46 orthogroups significantly expanded in the *Salicornia* lineage relative to other sampled Amaranthaceae species. The expanded families were enriched for functions related to photosystem maintenance, nutrient uptake and transport, secondary metabolism, protein homeostasis, and redox regulation—processes known to be critical for maintaining cellular function under saline and environmentally stressful conditions. These results are now presented in **Fig. 3a** and **Supplementary Fig. 7**.

Together, these new analyses substantially extend the functional exploitation of our genomic resources. Rather than revealing halophyte adaptation through expansion of canonical salt-stress transporter families, our results suggest that species in the *Salicornia* genus have achieved extreme salt tolerance through lineage-specific remodeling of conserved metabolic and stress-response pathways, a conclusion that is now explicitly discussed in the revised manuscript.

2. Given six high-quality *Salicornia* genomes, the lack of a thorough structural variation analysis is a major gap. The authors mention identifying three chromosomal translocations in the tetraploid *S. bigelovii* relative to diploids (validated by FISH), , with no gene losses at the breakpoints, but there is no systematic genome-wide SV survey. The manuscript should examine structural variants such as large insertions/deletions, inversions, copy-number variations, or presence/absence of genes across the species. Such analysis could uncover structural changes that contribute to species-specific adaptations (for example, gene duplications or deletions related to stress response). Currently, the focus on just a few translocations means potential SV impacts on salt tolerance or other traits remain unexplored. A comprehensive SV analysis across the genomes is needed to fully capitalize on the data.

Response: We agree with the reviewer that a more comprehensive analysis of structural variation (SV) was required to fully leverage the high-quality genome assemblies generated in this study. In response, we have now implemented a systematic, genome-wide SV analysis across all sampled *Salicornia* species.

For SV detection, we used *S. europaea* (Israel) as the reference genome for comparisons among diploid species. For tetraploid taxa, subgenome-resolved analyses were performed: the A subgenome of *Salicornia* sp. France and the B subgenome of *S. bigelovii* were used as references for SV

identification within each respective genome. This approach allowed us to account for subgenome divergence while enabling robust cross-species SV comparisons.

Our genome-wide analyses identified extensive structural variation across the genus, including insertions, deletions, and chromosomal rearrangements. Consistent with observations in other plant systems, the majority of SVs were located in intergenic regions (~79–84%). However, a substantial fraction overlapped gene bodies, including intronic (~11–16%) and exonic (~3–5%) regions. We therefore annotated genes overlapping SVs and performed Gene Ontology (GO) enrichment analyses to assess potential functional consequences.

In diploid *S. persica ssp. iranica*, which originates from a highly saline inland salt-lake environment (Supplementary Table 3), SV-overlapping genes were significantly enriched for stress-related functions, including “response to salt stress” (66 genes) and “response to osmotic stress” (72 genes). These findings suggest that structural variation may contribute to local adaptation by reshaping genes and pathways involved in osmotic and ionic stress responses. Similar analyses across other diploid species revealed both shared and species-specific enrichments related to ion transport, membrane-associated processes, and intracellular ion homeostasis.

These new analyses substantially extend our original findings beyond individual chromosomal translocations, providing a comprehensive view of structural genome variation across *Salicornia*. The results indicate that SVs contribute to genome architecture and may influence stress-response pathways relevant to halophytic adaptation, even in the absence of large-scale gene loss at rearrangement breakpoints.

The expanded SV methodology and analyses are now described in the Materials and Methods “*Structural variant and gene ontology enrichment analysis*” sub-heading and Supplementary Note 14. The results are presented under the sub-heading “*Comparative genomic analyses reveal extensive structural variation and gene family turnover in Salicornia*”. Genome-wide SV statistics are provided in Supplementary Table 7, gene–SV overlap summaries in **Supplementary Table 8** and GO enrichment results in **Supplementary Fig. 3c** and **Supplementary Tables 9–11**.

3. The population genomic analysis should be extended beyond basic clustering to identify genetic variation associated with salt tolerance or other adaptive traits. The authors present phylogenetic and structure analyses for the resequencing panel (Fig. 4 and Fig. 5), showing taxonomic groupings and some admixture. However, there is no attempt to connect these patterns to phenotypic or environmental variables. For a study centered on salt tolerance, it is crucial to ask: do certain populations or clades harbor alleles linked to higher salt tolerance? The manuscript should include analyses that probe adaptive divergence – for example, genome scans for differentiation between populations from high-salinity vs. lower-salinity environments, or association tests linking genomic variants to any available trait data. As is, the population analysis is purely descriptive (who is related to whom) and does not reveal any why or how regarding adaptation.

Response: We thank the reviewer for this insightful comment and agree that linking population genomic patterns to adaptive processes is critical for a study focussed on building halophyte resources for one of the key aims of understanding halophyte adaptation.

In response, we extended our analyses beyond population clustering to explicitly examine signatures of adaptive divergence where environmental contrasts and population sampling allowed. While many

Salicornia species in our resequencing panel are represented by accessions from geographically restricted locations—limiting the power of broad environmental association or genome-wide association analyses—two European tetraploid lineages provided a natural contrast for testing adaptive hypotheses. Specifically, *S. procumbens* accessions from the Netherlands and *Salicornia* sp. France originate from environments that differ markedly in soil salinity, seawater salinity, and seasonal temperature variability.

To focus on shared evolutionary variation, we conducted population genetic analyses on the A subgenome, calling variants on A-subgenome chromosomes of the French tetraploid assembly and comparing them with diploid relatives and *S. procumbens* (**Supplementary Fig. 14**). At an osmosensitive calcium-permeable channel (OSCA) locus on chromosome 6A, we detected contrasting Tajima's D patterns between the two tetraploid lineages. *S. procumbens* populations from cooler, low-salinity North Sea/Wadden Sea environments (soil EC 1.6–5.8 dS m⁻¹) exhibited strongly negative Tajima's D, consistent with a recent selective sweep. In contrast, *Salicornia* sp. France from the high-salinity, seasonally fluctuating Étang de Thau lagoon (soil EC ≈23.9 dS m⁻¹) showed positive Tajima's D, consistent with balancing selection (**Extended Data Fig. 10**). These patterns implicate soil-mediated osmotic stress as a driver of local adaptation and highlight OSCA-mediated Ca²⁺ signalling as a candidate adaptive pathway.

In addition, we conducted genome-wide analyses of nucleotide diversity (**Supplementary Fig. 17**) in *Salicornia* sp. France and *S. procumbens*, providing deeper insight into the distribution and filtering of genetic variation across contrasting saline environments. We further generated SNP density plots (**Extended Data Fig. 9; Supplementary Fig. 16a**) for species with sufficient population sampling to examine differences in genetic variation between A-genome diploid and AC-genome polyploid *Salicornia*.

For other taxa in the resequencing panel, adaptive analyses remain constrained by limited within-species sampling, strong interspecific population structure, and the absence of quantitative phenotypic data. We therefore revised the manuscript to explicitly frame our population genomic results as identifying lineage relationships, genomic diversity, and locus-specific signatures of selection, while emphasizing that expanded geographic sampling and phenotyping will be essential for future genome-wide association and environmental association studies.

These additions are now described in the Results under the sub-headings “*Genome-wide Nucleotide diversity (π) in AC subgenome Tetraploid*” and “*Detection of candidate selection regions using Tajima's D*”. The addition of a discussion relevant to these results can be found under the sub-heading “*Genetic diversity among diploid *Salicornia* species*”, with supporting analyses presented in **Extended Data Fig. 9, Extended Data Fig. 10, Supplementary Fig. 14, and Supplementary Fig. 17**.

4. The authors report population structure and admixture patterns in diploid *Salicornia* (e.g., mixed ancestry between *S. europaea* (Israel/Egypt) and *S. persica* ssp. *iranica* in Fig. 5a). However, they do not formally assess gene flow between species or populations. Given the misclassified accessions and close relationships, tests for introgression would be highly informative. The manuscript should include analysis of gene flow – for instance, D-statistics (ABBA-BABA tests) or other introgression methods – to determine if hybridization has occurred between certain lineages. This would add

evolutionary insight beyond the STRUCTURE analysis. Right now, the text infers admixture from population structure ($K=2-4$). Including such analyses would strengthen the conclusions about inter-species relationships and the evolutionary history of *Salicornia*.

Response: We thank the reviewer for this helpful suggestion and agree that formal tests of introgression provide important evolutionary insight beyond population structure analyses alone. In response, we conducted explicit tests for gene flow using D-statistics (ABBA-BABA tests) implemented in Dsuite across the Middle Eastern diploid *Salicornia* accessions.

These analyses reveal significant introgression among diploid *S. europaea* lineages, with particularly strong and statistically supported gene flow between accessions from Israel and Egypt (**Supplementary Table 19**). In contrast, introgression signals between *S. sinus-persica* (UAE) and *S. europaea* (Israel) were substantially weaker, consistent with long-term isolation and deeper lineage divergence. These results provide formal support for the admixture patterns inferred from population structure analyses (**Fig. 5a**), while clarifying that gene flow is uneven across lineages and largely restricted to geographically proximate *S. europaea* populations.

We have incorporated these findings into the Results under sub-heading “*SNPs genotyping population structure and gene flow in diploid Salicornia*” and in the Discussion under sub-heading “*Population structure, gene flow and phylogeny*”. We now explicitly distinguish between recent introgression among Mediterranean *S. europaea* populations and more limited gene flow involving *S. sinus-persica*. Together, these analyses strengthen our conclusions regarding interspecific relationships, lineage boundaries, and the evolutionary history of diploid *Salicornia*.

5. A major omission is the lack of analysis for selective sweeps or loci under natural selection related to salt stress. The study provides a wealth of genomic variation data, but the authors have not mined it for signals of adaptation. To merit publication in a high-impact venue, the authors should perform genome-wide selection scans (e.g., F_{ST} outlier analyses, XP-CLR, XP-EHH, or $\Delta\pi$ scans between groups). For example, if coastal vs. inland populations were sampled, are there regions of the genome with high differentiation that might indicate local adaptation to salinity? The text briefly notes F_{ST} values (e.g., $F_{ST} \approx 0.59$ between *S. europaea* (Israel) and *S. persica iranica*), but it does not go further to identify which genes or regions contribute to such divergence. The authors should identify any candidate genomic regions under selection and, importantly, highlight if those regions contain known abiotic stress response genes. As an illustration, if genes like HKT1;1 or NHX1 (known Na^+ transporters) or osmoprotectant synthesis genes show signatures of a selective sweep in high-salinity populations, that would directly link the genomic data to salt tolerance. Without such analysis, the paper misses an opportunity to connect genotype to adaptive phenotype.

Response: We thank the reviewer for this important suggestion. Our resequencing panel includes multiple *Salicornia* species sampled across broad geographic regions, most of which are obligate halophytes inhabiting highly saline environments. As a result, comparisons among species primarily reflect deep divergence, strong population structure, and limited gene flow, rather than recent local adaptation along clear environmental gradients (e.g., inland vs. coastal). For example, inland populations of *S. persica* ssp. *iranica* originate from salt lakes in the Al Qassim region of Saudi Arabia, where water ionic strength and osmotic conditions ($EC \sim 55 \text{ dS m}^{-1}$) were comparable to

coastal marine environments ($EC \sim 60 \text{ dS m}^{-1}$), at least at the time of collections. Under such conditions, genome-wide outlier-based methods (e.g., F_{st} scans) can be confounded by demographic history and divergence time, which may obscure locus-specific signatures of selection or generate false positives. Consistent with this, although pairwise F_{st} values were high (e.g., $F_{st} \approx 0.59$ between *S. europaea* (Israel) and *S. persica* ssp. *iranica*), genome-wide scans did not identify discrete, biologically interpretable candidate regions.

To improve detection power, we focused on closely related tetraploid lineages with shared genomic background but contrasting environments: *S. procumbens* [Netherlands, North Sea/Wadden Sea; low soil salinity ($EC \text{ 1.6--5.8 dS m}^{-1}$) and seawater salinity] and *Salicornia* sp. France [France, Étang de Thau lagoon; high soil salinity ($EC \text{ 23.9 dS m}^{-1}$) and fluctuating, higher seawater salinity]. Genome-wide Tajima's D scans identified a pronounced signal on chromosome 6A ($\sim 30 \text{ Mb}$) in the French lineage, overlapping a cluster of OSCA genes encoding osmosensitive Ca^{2+} -permeable channels associated with osmotic stress signalling (**Extended Data Fig. 10**). This region shows strongly positive Tajima's D in the French lineage, consistent with balancing selection in a highly variable saline lagoon environment, whereas the corresponding region in *S. procumbens* shows negative Tajima's D , consistent with a selective sweep under more stable environmental conditions.

These results provide direct evidence of adaptive divergence linked to salinity-associated osmotic stress pathways and identify OSCA gene as candidate loci underlying environmental adaptation in *Salicornia*. We have revised the manuscript to explicitly present these analyses and clarify the analytical framework, emphasizing locus-specific selection signals in appropriate phylogenetic and demographic contexts. The expanded analyses and interpretation are now described in the Results under "Detection of candidate selection regions using Tajima's D " and in the Discussion under "Standing genetic diversity and local adaptation in the A subgenome," with supporting data shown in **Extended Data Fig. 10**.

6. The RNA-seq differential expression analysis is not convincing in its current form. It appears the authors relied on a public transcriptome dataset rather than conducting a new, controlled experiment tailored to their hypotheses. As a result, the differential expression results are of limited scope and relevance. For instance, focusing on *Salicornia bigelovii* alone and using data not specifically designed to probe salt stress mechanisms makes it hard to draw firm conclusions. The manuscript states it identified genes associated with high shoot Na^+ accumulation via "comparative transcriptomic analyses" (abstract) but this was done by extracting DEGs unique to *S. bigelovii* from an existing dataset. This approach is suboptimal: factors other than salt tolerance could differ between *S. bigelovii* and others, and without a controlled experiment (same conditions, replicates, time points), the results are merely correlative. To strengthen this aspect, the authors should perform their own salt stress experiment or at least validate the key findings with additional data. The decision to highlight only two genes – SOS1 and SOS2 – in the transcriptomic analysis is not well-justified. These genes (components of the SOS salt tolerance pathway) are indeed important, but the exclusive focus on them suggests a very narrow analysis. The authors report that *S. bigelovii* uniquely upregulates SOS1 and has high baseline expression of SOS1 and SOS2 compared to quinoa or spinach. However, why limit the discussion to these two genes? There are many other salt-responsive genes (ion transporters, osmolyte synthases, LEA proteins, ROS-scavenging enzymes, transcription factors like DREB, bZIP, etc.) that could have been examined. If the intent was to validate the RNA-seq findings, focusing on only SOS1/2 gives an incomplete picture. The

manuscript would be far more compelling if a broader set of candidate genes from the DEGs were analyzed or at least if the authors explained that SOS1/2 were chosen a priori due to known importance. In addition, some validation of their expression (for example, qPCR in multiple species or conditions) would help. As is, the emphasis on just two genes appears arbitrary and insufficient, leaving the reader wondering if other crucial genes were overlooked.

Response: We thank the reviewer for this thoughtful and constructive critique. After careful consideration, we agree that the comparative transcriptomic analyses based on publicly available datasets were not sufficiently controlled to support strong conclusions about salt-stress mechanisms across species. In particular, differences in experimental design, growth conditions, developmental stage, and sampling strategy across studies limit the interpretability of differential expression patterns, and could lead to over-interpretation of correlative signals. In response to this concern, we have removed the comparative transcriptomic analyses derived from public datasets, including the identification of species-specific differentially expressed genes, the associated supplementary figures and tables, and any gene-level interpretations that relied solely on these cross-study comparisons. We therefore no longer present conclusions based on differential expression patterns inferred from external RNA-seq datasets. We acknowledge the reviewer's suggestion that new, controlled salt-stress experiments or cross-species expression validation (e.g. qPCR) would strengthen this aspect of the study. However, due to project funding and time constraints, it was not feasible to generate new RNA-seq datasets or perform targeted experimental validation within the revision period.

To ensure that the manuscript still meaningfully leverages the genomic resources generated here, we instead expanded the genome-based comparative analyses that do not rely on differential expression inference. Specifically, we conducted targeted gene-family *in-silico* analyses of multiple classes of genes previously implicated in salinity tolerance, including Na⁺/H⁺ exchangers (NHX), high-affinity K⁺ transporters (HKT), calcineurin B-like proteins (CBL), chloride channels (CLC) and slow anion channel-associated proteins (SLAC/SLAH). In addition, we performed genome-wide gene-family expansion and contraction analyses and a comprehensive structural-variation analysis. These results are described in detail in our responses to **Major Concerns 1 and 2** and are now fully integrated into the revised manuscript.

We have also revised the Abstract, Results, and Discussion to remove references to “comparative transcriptomic analyses” and to ensure that all remaining conclusions are supported by controlled genome-scale analyses presented in this study. We believe these revisions result in a more rigorous and appropriately cautious interpretation of the data, while still providing biological insight into the genomic basis of halophytic adaptation in *Salicornia*.

Minor concerns:

1. Which data did the authors use to generate the Ks value in Figure 3b? And it would be better to use the model fit to detect the peak.

Response: We thank you for raising these important questions. Ks values were calculated using the WGDI pipeline based on codon-aligned coding sequences. Specifically, the analysis used coding

DNA sequences representing the longest isoform per gene, gene annotation and genomic coordinate information to define gene models and syntenic relationships, gene length information required by WGDI for filtering, and protein homology data derived from BLASTP to identify homologous gene pairs. WGDI translated the coding sequences into proteins for alignment, inferred syntenic homologous pairs using protein similarity and genomic position, and then back translated the protein alignments to codon alignments from which Ks values were calculated. The corresponding contents are added in Materials and method sub-section “*Whole genome triploidization and divergence event analysis*”.

During preparation of the Ks values shown in **Fig. 3b**, we attempted Gaussian-mixture modelling to identify duplication peaks. The approach worked for tetraploid species, but failed for diploid species because the genome-wide Ks distribution contains a very large spike at $Ks \approx 0$ caused by abundant tandem duplicates. This near-zero mass dominated the fits and caused the mixture model to allocate components to the tandem signal rather than to broader paleopolyploid peaks, producing unstable and biologically implausible diploid/WGD components. For these reasons we did not rely on Gaussian-mixture decomposition for diploids. Instead, we present the empirical Ks histogram computed using the Nei & Gojobori (1986) [<https://doi.org/10.1093/oxfordjournals.molbev.a040410>] counting method (ks_NG86) with 350 bins, which more transparently displays the distributional features without introducing model-driven artefacts.

2.Line 265-266, I cannot find the evidence support this conclusion.

Response: We have added a read-depth plot for the diploid accession SalD_06, showing a 5A:5C translocation at the telomeric end of the short arm of the chromosome (**Extended Data Fig. 8d**). These patterns were uniform among A genome diploids.

3.Line 309, is it “bottleneck” or “founder effect”?

Response: *S. persica* ssp. *iranica* is the only inland population in our dataset and exhibits extremely low nucleotide diversity, indicative of a strong reduction in genetic variation. This pattern could result from either a historical bottleneck or a founder effect. Given its geographic isolation and likely derivation from coastal populations, a founder effect is plausible. Additionally, the allele frequency spectrum shows an excess of rare alleles, with very few alleles at intermediate or high minor allele frequency (MAF). This skewed distribution further supports the hypothesis of a recent founder event or strong population contraction. However, without explicit demographic modeling, we cannot definitively distinguish between a bottleneck and a founder effect. We have updated the manuscript accordingly.

4.How many single-copy gene groups have been identified in this study? Did the MCMCtree’s result have been tested by Tracer?

Response: We thank you for raising these important questions. In this study, we identified 1140 single-copy gene groups that were used for the downstream phylogenetic and divergence time analyses. The corresponding contents are added in the Materials and Methods in sub-heading “Phylogenetic analysis” section. Based on your question, the reliability of the divergence time estimation inferred by MCMCtree was further evaluated using Tracer. All key parameters showed

good convergence and sufficient effective sample sizes (ESS), indicating that the MCMCtree runs passed the convergence diagnostics. The detailed Tracer results are provided in **Supplementary Fig. 3**.

5. The authors write in the results that they used the IAGS to rebuild the ancestor genomes, but they did not write the detail methods in the “Methods” part. What about the evaluation results for this ancestor genome?

Response: Thank you for your question. Based on your comment, we have added a detailed description of the IAGS-based ancestral genome reconstruction in **Supplementary Note 14**.

To evaluate the reconstructed ancestral karyotype (**Supplementary Fig. 5**), we used the built-in evaluation module provided by IAGS, which calculates the completely rearranged endpoint ratio (CRB) and the corresponding estimated accuracy. In our analysis, the CRB was 0, indicating that no inferred endpoints were completely rearranged. Accordingly, the estimated accuracy of the reconstructed ancestral genome was 100.0%. It should be noted that this ancestral karyotype represents an inferred genome organization reconstructed from homologous protein-based syntenic blocks, rather than a real, extant genome.

6. the results in Extended Data Fig. 10, if the authors really want to explore something about the distribution pattern of SNP, they may separate the variations sets into different groups according to the Figure 5a. The authors indeed calculated the π and F_{ST} , which only provided very limited information, and where is the selection sweep analysis? I do believe that the authors could explain more based on their data.

Response: We thank the reviewer for this suggestion and for the opportunity to clarify and extend these analyses. First, we note that the figure referred to as **Extended Data Fig. 10** in the reviewer’s comments corresponds to **Extended Data Fig 9a–f** in the revised manuscript. In response to the reviewer’s suggestion, we have extended our SNP distribution patterns to other larger population groups with sufficient sample size (Extended Data Fig 9a–f), rather than treating all accessions as a single set. Specifically, we now present genome-wide SNP density and minor allele frequency distributions for:

- (i) all diploid *Salicornia* accessions combined,
- (ii) the largest single diploid group (*S. persica ssp. iranica*), and
- (iii) the A subgenome of tetraploid *Salicornia* sp. France (**Extended Data Fig. 9a–f**).
- (iv) the C subgenome of tetraploid *Salicornia* sp. France (**Supplementary Fig. 16**).

These analyses reveal pronounced heterogeneity in SNP density and allele-frequency spectra among groups, reflecting differences in demographic history, population size, and evolutionary context.

Regarding selective sweep analyses, we agree that such tests can be informative under appropriate demographic conditions. However, as discussed in our responses to **Major Comments 3–5** and now clarified in the manuscript, strong population structure, deep divergence among taxa, and limited within-population sampling inflate genome-wide differentiation metrics (e.g. F_{ST}) and obscure locus-

specific sweep signals. Under these conditions, standard genome-wide sweep scans are prone to false positives driven by demographic history rather than selection.

Accordingly, we focused our selection analyses on site-frequency–based statistics, Tajima's *D*, in closely related tetraploid lineages where demographic assumptions are more appropriate. This approach revealed a clear and localized signature of selection at the OSCA locus on chromosome 6A (Extended Data Fig. 10), as discussed in detail in The expanded analyses and interpretation are now described in the Results in sub-heading “*SNPs genotyping, population structure and gene flow in diploid Salicornia*” and Discussion sub-heading “*Standing genetic diversity and local adaptation in the A subgenome*”, with supporting data shown in Extended Data Fig. 9. Further information is also available in our responses to **Major Concerns 3 and 5**.

Together, the expanded SNP density and allele-frequency analyses, combined with targeted site-frequency–based tests in suitable population contexts, provide a more robust and interpretable framework for assessing patterns of genomic variation and selection in *Salicornia*.

7. In line 79-81, the authors said that there were several questions such as “evolution, gene flow and gene regulatory mechanisms”. But through this study, these questions still remain. Then what is the scientific questions and the main purpose for this paper?

Response: Thank you for this helpful comment highlighting the need to more clearly articulate the scientific scope and objectives of our study. We have substantially revised the Introduction to explicitly define the central scientific questions and aims of this work.

Accordingly, the aims of this study were five-fold:

- (i) to generate chromosome-scale reference genomes for representative diploid and tetraploid *Salicornia* species across major continents
- (ii) to resolve their evolutionary origins, polyploidization history, and subgenome architecture
- (iii) to characterise phylogenetic relationships, population structure, and genetic diversity across the genus using a globally representative resequencing panel
- (iv) to identify candidate genes associated with salinity-related traits through comparative genomics within *Salicornia* and across the *Amaranthaceae*
- (v) to standardize subgenome nomenclature and chromosome organization across published and newly generated *Salicornia* genomes, thereby preventing inconsistencies in public databases and ensuring reliable downstream comparative analyses. While questions relating to gene regulatory mechanisms and functional validation remain open, our study directly addresses long-standing uncertainties in *Salicornia* systematics, polyploid evolution, population structure, and genomic architecture that have previously constrained such investigations. We now explicitly frame these unresolved aspects as future research directions enabled by the genomic resources generated here, rather than as objectives fully addressed within the current study.

8.Line 87. What did the author mean for the “diversity”? What does the authors mean for “diversity assembled”?

Response: This sentence has been removed in the substantial revision of the Introduction. Now we provide a clearer description of previously assembled reference genomes and earlier phylogenetic efforts using molecular markers (L80-96).

9.Line 113-115. I don’t think this part belongs to the “Results”.

Response: We agree with the reviewer’s suggestion, and we have now relocated this sentence to the Introduction at L76-79 as well as clarified the sentences following.

10.The authors propose a mechanism wherein *Salicornia* (especially *S. bigelovii*) achieves extreme salt tolerance via constitutively high expression of Na⁺ transport genes like SOS1, enabling continuous sequestration of Na⁺ into vacuoles. While this is a plausible hypothesis – and is consistent with prior studies in *Salicornia*. the manuscript’s own data do not definitively prove this mechanism. There is no functional experiment (e.g., transgenic or mutant analysis) demonstrating that higher SOS1 or SOS2 expression causally leads to greater salt tolerance. Thus, any claims about “mechanisms” should be toned down.

Response: We agree with the reviewer that our data do not provide direct functional evidence demonstrating that elevated expression of Na⁺ transport components such as *SOS1* or *SOS2* encoding genes causally confers extreme salt tolerance in *Salicornia*. We have therefore revised the manuscript to avoid mechanistic overstatement and to more clearly distinguish hypothesis-generating observations from functional proof.

A key contribution of our prior research is the observation that the Na⁺/H⁺ antiporter SOS1 localises to the vacuolar membrane in *Salicornia bigelovii* (Salazar et al., 2024), whereas, at the time, SOS1 has classically been characterised as a plasma-membrane-localised transporter in glycophytic model species. While this finding is consistent with a role in intracellular Na⁺ sequestration and extreme salinity tolerance, we acknowledge that our current study does not include functional genetic experiments (e.g. mutants or transgenics) to establish causality.

At present, stable transformation and regeneration systems are not available for *Salicornia* species, which limits direct functional validation of candidate genes in their native context. To avoid speculative interpretation, we have removed cross-species transcriptomic comparisons and revised the text throughout the manuscript to frame our findings as associative and descriptive, rather than mechanistic. References to ‘SOS1 mechanisms’ have been removed.

We now explicitly state in the Discussion section under sub-heading “*Comparative genomics of salinity-related gene families*” that “Future functional validation of these candidate genes will involve the use of heterologous systems such as yeast (Salazar et al. 2024)[<https://doi.org/10.1038/s41467-024-48595-5>] as well as the development of *Salicornia*-specific genetic tools.”

11. The manuscript asserts that these genomic resources “lay the foundation for the domestication of Salicornia as a climate-resilient crop”. While the data (genome assemblies and a diversity panel) are undoubtedly valuable for crop development, this claim goes beyond what is demonstrated. The paper does not identify or improve any domestication traits (e.g., it does not discuss yield, seed shattering, growth rate, etc., which are critical for domestication). Claiming to pave the way for domestication is premature without showing how the genomic insights will be applied to breeding.

Response: We have revised our abstract to avoid unsubstantiated claims about how our research will contribute to domestication of Salicornia. However, we emphasize in both the abstract and the discussion in the sub-heading “*Genomic resources, breeding constraints and future crop development*” that well-curated genomic resources and germplasm collections form the backbone for future crop development.

Reviewer #1 (Remarks on code availability):

The computational analysis scripts are currently incomplete, which hinders reproducibility and review. Please provide the full, end-to-end code for each analysis to enable a thorough evaluation.

Response: We thank the reviewer for this important reminder. We have updated our GitHub repository to provide complete, end-to-end code and documentation for all computational analyses reported in the manuscript.

Reviewer #2 (Remarks to the Author):

The authors submitted a manuscript with the title "From Wild Halophyte to Future Crop: Genomic Insights into Salicornia". The study examines the genomes of several Salicornia species. As part of the Amaranthaceae family, Salicornia is appreciated for its edible succulent stems and oil-rich seeds, though enhancement of its genetics has been challenged by limited genomic data. Chromosome-scale, annotated genome assemblies for six Salicornia species have been provided, with four distinct subgenomes identified within the genus. Genes linked to high shoot Na⁺ accumulation, a crucial mechanism for salt tolerance, have been discovered through comparative transcriptomic analyses. Interspecies relationships and misclassification at both accession and species levels are illuminated by phylogenetic clustering and population structure analyses using a globally representative resequencing panel. With these insights, a solid basis is formed for Salicornia's development as a climate-resilient crop tailored for saltwater agriculture.

The sequencing of six Salicornia genomes is indeed excellent news for the scientific community, offering a wealth of data that will hopefully soon be publicly accessible. However, a primary concern is that, while the manuscript presents a significant amount of data, the information is not thoroughly interconnected, evaluated, or deeply interpreted. This omission results in several major and minor contradictions throughout the manuscript. It's important to recognize that journals like Nature Communications have space constraints that limit the extent to which results can be elaborated in

detail. Nevertheless, the authors have a crucial responsibility to strategically focus on and highlight the most significant findings while also addressing and not overlooking any contradictory information. By improving the integration, evaluation, and interpretation of the collected data, the manuscript can effectively convey its contributions to the field and maintain a coherent narrative. This approach will ensure that the results are presented with clarity and precision, thus enhancing the overall impact of the study.

Note: Sequence of authors in the invitation email/submission email is not the same as in the paper itself with respect to the author Vanessa Melino.

Abstract and title

The current title of the manuscript is somewhat cryptic and could be more descriptive. To enhance clarity, consider revising it to better reflect the focus on genome data and its implications for developing *Salicornia* spp. as a crop plant. While the manuscript provides valuable insights into genome data, which is certainly of high interest for advancing *Salicornia* as a crop, it lacks concrete facts on its development as a crop plant. To address this gap, consider integrating more detailed information pertinent to crop development. Additionally, the abstract could be strengthened by incorporating more quantitative data. At present, it is quite general, but could be elevated by including key numbers and precise information. Look to abstracts from other sequencing studies as examples, which often highlight quantitative findings to convey the significance of their research upfront. By implementing these revisions, the manuscript would not only become more accessible and informative but also underscore its relevance to the advancement of *Salicornia* spp. as a crop species.

Response: We thank the reviewer for this thoughtful and constructive assessment of our manuscript. We appreciate the recognition of the value of the six *Salicornia* genome assemblies and fully agree that, for a journal such as *Nature Communications*, it is essential to present a coherent, integrated narrative that highlights the most significant findings while explicitly addressing apparent contradictions or limitations.

In response to this overarching concern, we have substantially revised the manuscript to improve integration, evaluation, and interpretation of the datasets. Specifically, we have refocused the Results and Discussion to emphasize the key biological insights enabled by these genomic resources, rather than presenting datasets in parallel. The revised manuscript now provides a critical synthesis of comparative genomics (including genome structure, subgenome architecture, repeat dynamics, structural variation, and gene family expansion/contraction), population genomics (phylogeny, population structure, introgression tests, nucleotide diversity, and Tajima's D), and cytogenetics (karyotype validation by FISH). Where interpretations are associative rather than mechanistic, this is now stated explicitly.

In line with the reviewer's comments on scope and framing, we have revised the title to more accurately reflect both the scale of the genomic resources and the biological questions addressed:

“Chromosome-scale genomes and a resequencing panel of *Salicornia* resolve subgenome diversity, species boundaries and signatures of halophyte adaptation.”

The revised title avoids over-emphasizing immediate crop development and instead highlights the manuscript's core contributions: the establishment of a standardized, phylogenetically informed genomic framework for *Salicornia*, the curation of a genetically validated global germplasm panel,

and comparative analyses that identify loci and genomic features potentially contributing to halophyte adaptation.

We have also strengthened the abstract by reducing general statements and incorporating more concrete, quantitative information (e.g. number of genomes, subgenomes resolved, population scale, and types of analyses performed), following conventions of other genome-focused studies in Nature Communications. At the same time, we have moderated statements related to crop domestication to ensure they are appropriately forward-looking and supported by the data presented.

Regarding the reviewer's note on author order, we confirm that Dr Vanessa Melino is the lead and corresponding author of this study. An incorrect author sequence appeared in the submission invitation system due to a technical entry error during registration. This issue was reported to Nature Communications on the day of submission, and we were advised that the author order listed in the submitted manuscript—which is correct—will be used for publication.

Overall, we believe these revisions substantially improve the clarity, coherence, and impact of the manuscript, ensuring that the major findings are clearly articulated, internally consistent, and aligned with the expectations of Nature Communications.

Introduction

The main part provides a good introduction into the topic. Some general statements are taken unquestioned from other studies, some references are certainly missing (see later in this review). For this type of journal, Nature Communications, the formulation of biological hypotheses is necessary, at least from my point of view. Otherwise, it remains a descriptive sequencing paper that could be published in more suitable journals. Although the approach and the data produced are certainly of high interest for future studies.

Thank you for this valuable suggestion, we now provide a more detailed description of the knowledge gap that our research will address as well as the aims. This helps define the biological questions that we address with this research.

Response: We thank the reviewer for this important comment regarding the need for clearly articulated biological questions and hypotheses, particularly for a journal such as Nature Communications. We agree that without this framing, the work risks being perceived as primarily descriptive, despite the scale and depth of the genomic resources generated.

In response, we have substantially revised the Introduction to more explicitly define the knowledge gaps, biological motivation, and questions addressed by this study. The revised Introduction now moves beyond general background to clearly articulate why *Salicornia* represents a compelling system for studying extreme salinity adaptation and what has been missing that has prevented progress in this area.

Specifically, we now emphasize that despite *Salicornia*'s exceptional physiological salt tolerance and its relevance for saltwater-based agriculture, fundamental questions remain unresolved regarding:

1. the evolutionary origins and polyploidization history of the genus,
2. the identity and structure of subgenomes in tetraploid lineages,
3. patterns of gene flow, population structure, and taxonomic boundaries across continents, and
4. whether genomic changes such as structural variation, copy-number variation, and gene-family remodeling are associated with adaptation to extreme saline environments.

5. A framework that standardizes subgenome designation and chromosome naming of publically available and our own contributed assembled genomes.

To ensure these questions are explicit, we have reframed the Introduction around five clearly stated aims (L113-121), which now serve as testable, biologically motivated objectives rather than purely technical goals. These aims connect global challenges in water security and salinization to evolutionary genomics and adaptation, positioning *Salicornia* not simply as a sequencing target, but as a model system for understanding extreme halophyte adaptation and a foundational resource for future crop development.

In addition, the revised Introduction explicitly acknowledges where hypotheses are associative rather than mechanistic, setting appropriate expectations for the scope of inference supported by the data. This framing aligns the study with the standards of Nature Communications, where genome-scale resources are expected to address clear biological questions while enabling future functional work.

We believe these revisions substantially strengthen the conceptual foundation of the manuscript and clarify its scientific purpose, moving it from a descriptive sequencing effort to a hypothesis-driven, integrative genomic study of halophyte evolution and adaptation.

Materials and Methods

Several preprints are cited that have not yet been peer-reviewed, they might never be fully published. Some details are missing (e.g., conditions in the greenhouse). The homogeneity of seed material is not mentioned. Therefore, it is important to mention the number of plants that were used/pooled for the analyses, for example for the RNA extraction.

In the case of the SNP data, large quantities were apparently collected, which later could scarcely be explored in detail.

As I am not an expert in bioinformatics, I am not commenting on the methods evaluating the sequencing data.

Response: We thank the reviewer for their careful reading of the Materials and Methods section and for highlighting issues related to reproducibility, transparency, and data usage.

First, with respect to cited preprints: the previously referenced laboratory protocols that were available as preprints at the time of initial submission have now been peer-reviewed and formally published, and the reference list has been updated accordingly.

Second, regarding greenhouse conditions and biological replication: we have clarified growth conditions and experimental design in the revised Methods. RNA was isolated from individual plants or individual organs collected from a single plant ($n = 1$ biological replicate) and used for Iso-seq or RNA-seq solely to support genome annotation, including gene model prediction and transcript structure validation. For this purpose, biological replicates are not required. We have clarified the replicate number in the Materials and Methods.

Third, with respect to seed material homogeneity: we now explicitly state that accessions were derived from single-plant collections maintained as selfed lines. This clarification has been added to the germplasm and sampling description.

Fourth, regarding the SNP data: in the revised manuscript, we have substantially expanded our population genomic analyses and now leverage the resequencing dataset more fully. Specifically, we focus on questions directly aligned with the study aims, including refined phylogenetic, inference using both SNP and k-mer-based framework appropriate for mixed ploidy, formal tests of introgression using D-statistics, genome-wide diversity metrics (π , SNP density and minor allele frequency), and locus-specific signatures of selection where population sampling permitted (e.g. Tajima's D at candidate loci). These additions move the analysis beyond descriptive clustering and provide explicit evolutionary and adaptive context. We also clarify in the revised text that this large SNP dataset represents a foundational community resource, enabling future hypothesis-driven analyses with increased sample sizes—such as environment–genotype associations or trait mapping.

Overall, these revisions substantially strengthen the rigor, clarity, and focus of the Materials and Methods section and ensure that all analyses presented are well supported by the underlying experimental design and data.

Results

The authors collected a lot of interesting and relevant data. Some results as the ploidy levels might be so important for this particular study to present them in the main part. For example, in the appendix it is shown that the 2C genome size of *Salicornia europaea* (France) does not correspond to the known genome sizes of *S. europaea*. This difference is also visible in the phylogenetic tree. Therefore, the sequenced genome should better not be labeled as *S. europaea* in contradiction to the literature mentioned earlier. It could indicate to another species and a few clear traits exist for species identification, aside from the ETS marker sequences (Kadereit et al. 2012) and genome size (Zonneveld 2019). In lines 227-229, it states: "*S. procumbens* accessions formed a distinct clade within the larger tetraploid *S. europaea* clade from France (Fig. 4 and Extended Data Fig. 7)." One could ask why the species from France was not classified as *S. procumbens*?

Response: Thank you for your constructive comments. We agree that ploidy levels are important for interpreting our results. We have presented the ploidy data in Figure 4. We agree that a tetraploid *Salicornia europaea* has not been prior described and the Eurasian tetraploids should fall under the *Salicornia dolichostachya* clade, according to Kadareit et al. 2007

[<https://doi.org/10.2307/25065909>]. We originally assigned the name *S. europaea* L. in this manuscript to describe the specimens collected in France and the accession used to assemble a reference genome based on the nomenclature found in the French administrative and biodiversity databases, consistent with the national TAXREF nomenclature (INPN CD_NOM identifier 119881).

However, we agree that multiple lines of evidence indicate that this designation is not appropriate. First, flow cytometry revealed a 2C genome size for the French accessions that does not correspond to known genome sizes of *S. europaea* (Zonneveld, 2019)[<https://doi.org/10.1111/j.1438-8677.2011.00522.x>]. Second, phylogenetic analyses and PCA place these French tetraploids in a distinct subclade closely related to, but clearly separated from, *S. procumbens* (Fig. 4; Extended Data Fig. 7 and Supplementary Fig. 13), addressing the reviewer's question as to why they were not classified as *S. procumbens*. Although the two groups share the same subgenomic composition, they show consistent morphological differences and form separate phylogenetic lineages, making species-level assignment uncertain.

To further clarify their genomic composition, we mapped Illumina short reads from the French tetraploid accessions and other accessions to the chromosome-scale T2T tetraploid *Salicornia* reference genome from China (Duan et al., 2025), which they identified as tetraploid *S. europaea*. Reads from the French tetraploid aligned almost exclusively to the A subgenome, with negligible alignment to the B subgenome (Supplementary Fig. 15a,b; Supplementary Table 18) of the published genome. In addition, diploid A-genome accessions showed ~80% alignment to the A subgenome and none to the B subgenome. These results indicate that the French tetraploid does not correspond to the published AABB tetraploid reference (Duan et al. 2025)[<https://doi.org/10.1093/plphys/kiaf565>] and instead harbors a distinct, previously uncharacterized subgenome ‘C’. The results are described under sub-heading “*Nomenclature of Salicornia sp. France and relationship with published S. europaea tetraploids from China*”.

Based on this combined evidence—genome size, phylogenetic placement, morphology, and subgenomic composition—and following discussion with germplasm curators, we have revised the manuscript to refer to these accessions as *Salicornia* sp. France. We now explicitly present this tetraploid as a distinct, currently unnamed species that provides a valuable genomic resource, including a novel C subgenome not represented in the published tetraploid reference.

We have therefore adopted the provisional designation *Salicornia* sp. France to denote a taxonomically unresolved lineage. Formal species assignment will require targeted taxonomic revision integrating genomic, cytogenetic, and morphological expertise.

What was the reasoning to select these species for the phylogenetic tree, please add some information.

Response: Thank you for the suggestion — we agree and have added a short explanatory paragraph to the Supplementary Note 12 clarifying our taxon selection.

Briefly, the added text explains that *Arabidopsis thaliana* was used as an outgroup because of its well-annotated reference genome and position within core eudicots; non-halophytic/weakly halophytic relatives (e.g. *Amaranthus*, *Beta*, and three *Spinacia* species) were included to provide family-level phylogenetic context and reduce annotation bias — the three *Spinacia* taxa were chosen because high-quality assemblies/annotations are available and they represent both cultivated (*Spinacia oleracea*) and wild (*Spinacia tetrandra*, *Spinacia turkestanica*) lineages; multiple independent halophytic lineages (*Atriplex*, *Haloxylon*, *Chenopodium* and *Suaeda* subgenomes) were included to compare salt-adapted clades and to explicitly account for confounding factors such as polyploidy, lineage-specific gene-family expansions, and copy-number variation in transport-related genes.

Line 93-95 explicitly criticizes that many species have been labeled as *S. europaea* at various locations. Line 122 mentions the supposedly tetraploid *S. europaea* species, which according to Kadereit et al. 2012 cannot exist, as only the genetically similar *S. perennans* species include tetraploid types; moreover, *S. procumbens* is another known European tetraploid species. The situation is unclear because Table 1 indicates a 2C genome size of 1129 Mb (which corresponds to *S. perennans* in Zonneveld 2019), but the measured pg content isn't mentioned, which should have been 2.309 pg correspondingly, suggesting the tetraploid *S. procumbens* species. The fact that speaks

against *S. procumbens*, as mentioned earlier, is that *S. europaea* (France) contains somewhat similar subgenomes compared to *S. europaea* (and other diploid species), indicating an assignment to the diploid *S. perennans* species, though the latter isn't mentioned either. While I somewhat understand the line of reasoning, I can't draw clear conclusions besides noting these contradictions due to the missing pg content.

Response: We thank the reviewer for highlighting these contradictions and gaps in our research. We now refer to tetraploid *Salicornia europaea* (France) as *Salicornia* sp. France, and for the reasons described below, we cannot assign this species to other characterised European tetraploid species mentioned by the reviewer. We have completed additional experiments to establish a robust karyotype for tetraploid *Salicornia* sp. France including additional oligo painting to solve a discrepancy between DAPI stained chromosome counts and the genome assembly size (Extended data Fig. 4). We now also include the measure pg of nuclear DNA content in **Supplementary Table 3** for all samples assessed by flow cytometry. It is worth noting that *Salicornia procumbens* has a slightly larger genome size (1220 Mb, **Supplementary Table 3**) than that of *Salicornia* sp. France (1129 Mb, **Table 1**), however, both karyotypes have been resolved as $2n = 4x = 36$ (**Extended data Fig. 2 and Extended data Fig. 4**).

Importantly, genome-wide read depth coverage analyses based on Illumina short-read mapping (e.g., Extended Data Fig. 8b) reveal that the French tetraploid possesses a unique AC subgenomic composition, whereas diploid taxa—including *S. europaea* and other diploid species—carry only the A subgenome. Although the French tetraploid is phylogenetically closest to the European tetraploid *S. procumbens* (Fig. 4), the two form distinct subclades and are separable by consistent morphological differences. These genomic and phenotypic distinctions, together with differences in genome size, preclude confident assignment of the French tetraploid to either *S. procumbens* or *S. perennans*.

Response: Based on this integrated evidence—flow cytometry with pg measurements, cytogenetics, genome size, subgenomic composition, phylogenetic placement, and morphology—and following discussion with the original collectors and germplasm curators, we designate these accessions as *Salicornia* sp. France. We have revised the manuscript accordingly and now describe this reasoning in detail in the Results subsection titled “*Nomenclature of Salicornia* sp. France and relationships with published *S. europaea* tetraploids from China.”

When referring to *Salicornia* as an excellent protein source, it is surely the seeds being referred to, which, to my knowledge, should not be consumed in large quantities due to the saponins they contain. (Line 66).

Response: Yes, we are referring to the seed quality and we have now clarified these sentences to be more explicit L65-68: “*Salicornia* L., (Amaranthaceae; Salicornioideae) commonly referred to as glasswort or samphire, comprises succulent halophytes cultivated for their edible shoots (sea asparagus), and for seeds that are rich in nutrients, proteins and oils that are an excellent source of fats.”

There is no clear distinction between *Salicornia*, *Sarcocornia*, and *Tectocornia*. It should be mentioned that these are different genera (Line 68).

Response: Thank you for this comment, we have distinguished the genera more clearly now in our introduction lines 68-71: “Most annual *Salicornia* species are widely distributed across four continents⁷, while related genera show more restricted distributions, including *Sarcocornia* (predominantly perennial, with a broad intercontinental range) and *Tecticornia* (comprising both annual and perennial species, largely endemic to Australia).”

Lines 273-276: This should be part of the discussion.

"Smaller collections of *S. europaea* from China and the Netherlands also formed a separate clade, clustering with *S. brachiata* (Fig. 4)." So, are Dutch, Indian, and Chinese species closely related? I find no conclusions from this result.

Response: We thank the reviewer for this comment and agree that the statement in Lines 273–276 is more appropriately placed in the Discussion. We have therefore moved this text accordingly.

Regarding the apparent clustering of *Salicornia* accessions from China, the Netherlands, and India (*S. brachiata*), we caution against over-interpretation of this pattern. Our resequencing panel includes multiple *Salicornia* species; however, for several taxa the sample size is very limited. For example, *S. brachiata* is represented by only four accessions. Such small sample sizes substantially reduce statistical power and prevent robust inference about phylogenetic relationships or historical connectivity among these populations.

While these accessions form distinct subclades within a broader clade (Fig. 4), this pattern primarily reflects their divergence from the Middle Eastern *S. europaea* group (Israel, Egypt, and UAE) rather than evidence for close evolutionary relationships among Dutch, Indian, and Chinese populations. We therefore treat this observation as descriptive and refrain from drawing broader biogeographic or evolutionary conclusions, which would require denser sampling across regions and species.

Line 412: "Cross-species hybridization has historically been difficult." References?

Response: We thank the reviewer for this suggestion. We have now included a citation to Kadereit *et al.* (2007) [<https://doi.org/10.2307/25065909>], which documents extensive reproductive isolation and limited crossability among *Salicornia* lineages. This observation is also consistent with our own experimental experience, as now noted in the revised manuscript Discussion in sub-heading “*Genomic resources, breeding constraints and future crop development*”.

How do sequenced genomes help overcome cross-breeding incompatibility? If anything, more data has been gathered on why certain species are not crossable (Line 414).

Response: We agree with the reviewer. High-quality reference genomes provide a framework for understanding the genetic and chromosomal bases of reproductive isolation, including ploidy differences, karyotypic rearrangements, and sequence divergence that underlie cross-breeding incompatibility. In the case of *Salicornia*, the availability of chromosome-scale assemblies and population genomic resources enables precise identification of subgenome structure, chromosomal

collinearity, and structural variation, all of which inform why particular species pairs are not crossable. Importantly, these genomic insights also provide alternative routes for trait introgression. We have revised the manuscript to clarify that our genomic resources facilitate *understanding* and *circumventing* genetic barriers to crossing, rather than directly overcoming reproductive incompatibility. This is updated in Discussion under sub-heading “*Genomic resources, breeding constraints and future crop development.*”

Lines 417-429: All the effort of intra-family transcriptome comparison just to bring up an already known pathway (SOS) and specifically the SOS1 transporter, which has already been frequently discussed in connection with *Salicornia*? The potential eco-physiological adaptations (larger vacuoles; habitat specialization) due to ploidy levels (Kaligarić et al. 2008) were not discussed, even though collecting wild types provided an opportunity for closer examination: "We could add that annual tetraploids, derived from diploids, then occupied the even more extreme, previously empty space of mudflats along shallow coasts, where tidal activity is a limiting growing factor, with its periodicity of flooding." Already Noble et al. (1992) report that tetraploids tend to be more characteristic for the lower parts of the marshes." Why did you select *S. biglovii* for more detailed gene expression analysis? A comparison of all six species would have been interesting. The authors could think about hybridization events to explain some of the data.

Response: We thank the reviewer for this insightful comment. Given the limited direct evidence available, we have removed the final intra-family transcriptome comparison section to avoid over-interpretation of conserved salt-stress pathways. Instead, we now focus on a more conservative comparison between the two tetraploid species, *Salicornia* sp. (France) and *S. procumbens*, which occupy distinct habitats, and discuss differences in their genetic profiles in relation to environmental context. We have incorporated these findings into the Results under sub-heading “*Detection of candidate selection regions using Tajima’s D*” and in the Discussion under sub-heading “*Detection of candidate selection regions using Tajima’s D*”.

Based on the available data from both tetraploid and diploid *Salicornia*, we find that habitat differentiation alone does not allow a clear inference of eco-physiological adaptations associated with ploidy level in *Salicornia*. Accordingly, we have confined our interpretations to conclusions that are robustly supported by the existing data. We acknowledge that a comprehensive comparison of salt-responsive gene expression across all six species would be highly informative; however, due to the conclusion of project funding and space limitations, we were unable to conduct the additional growth and transcriptomic experiments required for such an analysis.

There are some more examples like these. Therefore, a carefully revision of each sentence to draw final (or prefinal) conclusion based on a combination of the various data sets is highly recommended.

We thank the reviewer for highlighting the importance of clearer integration across the diverse datasets generated in this study. In the revised manuscript, we have made explicit how conclusions are drawn from the combined interpretation of multiple, independent genomic and population-genetic analyses, rather than from any single dataset in isolation.

Specifically, our conclusions on *Salicornia* evolution and genome architecture are supported by the concordance between chromosome-scale genome assemblies, cytogenetic validation (oligo-FISH), k-mer-based phylogenetic clustering, subgenome assignment using sequence composition, and comparative synteny analyses. These complementary approaches independently resolve tetraploid structure and subgenome composition and together provide a robust framework for inferring polyploid origin and subgenome evolution.

Similarly, insights into population structure, species boundaries, and gene flow arise from the integration of k-mer phylogenies, SNP-based PCA, population-structure analyses, pairwise *F*_{st} estimates, and formal tests of introgression (D-statistics). The consistency of signals across these methods underpins our conclusions regarding misclassified accessions, regional differentiation, and asymmetric gene flow among Middle Eastern diploid lineages.

Finally, evidence for local adaptation and lineage-specific evolutionary pressures is derived from the combined analysis of genome-wide diversity metrics (π), allele-frequency spectra, Tajima's D scans, environmental metadata, and comparative genomics of stress-associated gene families. The convergence of population-genetic signatures with ecological context and gene-family remodeling strengthens the biological interpretation beyond descriptive patterns.

To improve clarity, we have revised the Discussion to explicitly articulate these linkages between datasets, ensuring that each major conclusion is grounded in multiple, independent lines of evidence. This integrative framework is central to the study and distinguishes it from purely descriptive genome reports.

Discussion

A lot of data have been collected, but only few are discussed. The first paragraph of the discussion can be deleted as everything was written in the introduction already. That gives some space for a deeper discussion. The biological results are only very shortly discussed. Many relevant comparable facts have been already published and that means several references are missing.

Response: We thank the reviewer for this constructive feedback on the scope and depth of the Discussion. In response, we have made substantial revisions to strengthen the biological interpretation and contextualization of our results.

First, as suggested, we deleted the opening paragraph of the Discussion, which largely reiterated background material already covered in the Introduction. This created space to focus more explicitly on interpretation of the data.

Second, we restructured the Discussion using thematic subheadings (e.g. polyploid origin and subgenome evolution; standing genetic diversity and local adaptation; population structure and gene flow; conservation implications; breeding constraints and future crop development), allowing us to engage more deeply with the major biological findings emerging from the genomic and population-scale analyses.

Third, we have expanded the biological interpretation of key results, including (i) the evolutionary implications of resolving four distinct *Salicornia* subgenomes and the absence of extant progenitors for B and D, (ii) contrasting demographic and selective regimes acting on the A subgenome across saline environments, and (iii) the relevance of population structure and introgression patterns for species boundaries, conservation, and breeding.

Finally, we have substantially increased citation of relevant prior work throughout the Discussion, including comparative examples from halophytes, polyploid crop relatives, and Amaranthaceae genomics, to better situate our findings within the existing literature and acknowledge comparable published observations.

Together, these changes ensure that the Discussion now moves beyond description, provides a deeper synthesis of the biological insights enabled by the dataset, and clearly articulates how our results advance understanding of halophyte evolution, adaptation, and applied potential.

What is the exact meaning of subgenomes? Can you draw your conclusions based on the six genomes? I assume that more data are needed to prove your hypothesis.

It is highly recommended to discuss your data that means to compare to other genera or families.

Response: By “subgenome” we refer to one complete set of homoeologous chromosomes within an allopolyploid, each derived from a distinct ancestral diploid lineage. In the *Salicornia* allotetraploids analysed here, the 18 chromosomes can be partitioned into two such subgenomes (e.g. A and C, or B and D), each comprising nine chromosomes that show strongest synteny and sequence similarity to one of the diploid reference genomes (with the exception that no diploids with the C genome were identified in this study). We have clarified this definition and the criteria used for subgenome assignment in the Results sub-section “*Chromosome-scale genome assemblies and annotation quality*”.

We agree that additional genomes from other Salicornioideae genera would further refine the deeper evolutionary history of the group. Our main conclusions, however, are directly supported by the six chromosome-scale reference genomes representing the major diploid and tetraploid *Salicornia* lineages, together with a global resequencing panel. We have revised the Discussion section subheading “Polyploid origin and subgenome evolution in *Salicornia*” to clarify that our inferences about polyploid origin and subgenome structure apply to the sampled lineages, and we do explicitly state that broader taxon sampling will be required to test whether this model holds across the entire subfamily.

We compared the *Salicornia* assemblies generated in this study with quinoa (*Chenopodium quinoa*), as described by Rey et al. (2023) [<https://www.nature.com/articles/s41597-024-04200-4>], using sequence similarity analysis with minimap, but observed very poor alignment, likely reflecting the substantial divergence between the species. We also compared BUSCO ortholog copy sets between quinoa and the published *S. depressa* genome, which the authors reported using for chromosome assignment. However, we found that the chromosome and subgenome assignments in *S. depressa* were not consistent with those of quinoa reported by Rey et al. (2023). Following their chromosome assignment strategy, we extracted BUSCO ortholog copy sets from quinoa and identified corresponding copies in *S. depressa*. Nevertheless, discrepancies in chromosome numbering and

orientation between *Chenopodium* and *S. depressa* remained. We have noted these inconsistencies and communicated them to the authors.

Lines 394 – 398: Actually, this is known already and could be deduced from previous data.

Response: Thank you for this comment. Lines 418-420 of the original manuscript refer to the documented loss and degradation of saline habitats supporting *Salicornia* populations. We agree that habitat loss and pressure on saline ecosystems have been noted previously, and our intention was not to present this as a novel finding. Rather, we included this context to situate our genomic results within the broader ecological and conservation framework relevant to *Salicornia*.

To clarify this point and avoid any implication of novelty, we have revised the text to more explicitly acknowledge that these patterns have been reported in prior studies and to frame our discussion as building upon, rather than newly deducing, this existing knowledge.

Lines 418-420: As you are describing gene expression, please write gene names in Italics.

Response: This sentence was structured so that SOS1 referred to the encoded protein, however, this is no longer in the substantially revised version of the manuscript.

Line 427: Salt Overly Sensitive SOS pathway: The abbreviation is used already several times and now it is introduced, please correct.

Response: Thank you, this sentence is no longer present in the substantially revised version of the manuscript.

Minor

There are many writing and formatting errors that need to be very carefully corrected, too many to point them out individually.

Response: Thank you for your comment, we have now reviewed the manuscript in detail and found small grammatical/spelling errors that have been corrected. We have also paid particular attention to formatting SI units, hyphenation, punctuation and italicised species latin names.

Also the list of references contains many writing and formatting and fact (wrong years) errors.

Response: Thank you for noting this. We agree, and we have now thoroughly checked and corrected the reference list, including formatting consistency, citation completeness, and factual errors such as incorrect publication years. We also updated any references that were previously cited as preprints where peer-reviewed versions are now available, and ensured that in-text citations and the reference list are fully concordant throughout the revised manuscript.

The Bioproject data were not yet available but these will be hopefully available with the acceptance of the manuscript.

Response: Thank you for raising this point. The BioProject data are not yet publicly available, but will be released immediately upon acceptance of the manuscript (see Data Availability in main manuscript). In the meantime, the genome assemblies and annotation files are available for review via the Dryad reviewer link that we have provided to reviewers both in the manuscript text (http://datadryad.org/share/43tnCfOleGD2ThzJ9ngKwGNID8_TbKj7Yz2HE1kl3sE) and in the data access document titled “Data Access.pdf”.

I personally, do not recommend to write in the first person but in a passive way.

Response: Although we understand your point of view, Nature communications writing guidelines strongly encourage use of the active voice and allows the first person “we” throughout the manuscript. They do however, prefer a passive voice when method-focussed so we have taken your advice to ensure the the passive voice is used through-out the methods.

Reviewer #3 (Remarks to the Author):

The paper mainly reports six genomes of *Salicornia* species and a basic population genetics analysis that partially resolves relationships within the genus. Despite the ambitious title “From Wild Halophyte to Future Crop: Genomic Insights into *Salicornia*”, the paper does not demonstrate how these resources will accelerate *Salicornia*’s domestication or improve agronomic traits. Moreover, the genomic story lacks depth and novelty, leaving significant opportunities to strengthen both its scientific impact and its relevance to genome evolution.

Response: In response, we have substantially expanded and integrated the comparative genomic analyses across *Salicornia* and related Amaranthaceae species. Specifically, we now present genome-wide structural variation (SV) analyses (**Fig. 2c; Supplementary Tables 7–11**), subgenome-resolved copy-number variation, and gene family expansion and contraction analyses (**Fig. 3a; Supplementary Figs. 6–7; Supplementary Table 12**), which together reveal lineage-specific genomic remodeling associated with adaptation to saline environments. Notably, we identify *Salicornia*-specific expansion of chalcone synthase (CHS/TT4) genes and convergent expansion of Cupin_1 gene families across high- Na^+ -accumulating halophytes, placing these changes within an explicit evolutionary and comparative framework. In addition, we performed targeted comparative analyses of ion transport– and stress-associated gene families (NHX, HKT, SLAC/SLAH, CBL, CLC; **Supplementary Figs. 8–12; Supplementary Tables 13–15**), demonstrating that extreme salt tolerance in *Salicornia* is not driven by large-scale expansion of canonical transporter families, but rather by selective reorganization and amplification of conserved stress-response pathways. These results place our findings in an explicit evolutionary and comparative context and substantially strengthen both the depth and novelty of the genomic narrative.

We have also revised the title and abstract to more accurately reflect the scope and contributions of the study, emphasizing genomic architecture, evolutionary history, and adaptive diversity, rather than direct claims of domestication outcomes. While we do not claim to deliver agronomic trait improvement in this study, the genomic resources (**Table 1**), subgenome resolution (**Fig. 2a–b**), and population-scale analyses (**Figs. 4–5; Supplementary Table 19**) we provide establish a necessary foundation for future functional and breeding efforts.

Collectively, these revisions substantially enhance the depth, novelty, and evolutionary insight of the manuscript and clarify its contribution to genome evolution and halophyte adaptation.

More comments are listed below.

Line 83-87. The authors criticize earlier *Salicornia* genomic studies for their low assembly quality and incomplete annotations. The authors should provide concrete evidence to support this claim. Additionally, there is a misreferenced citation.

Response: Thank you for this constructive comment. We agree that statements regarding the quality and completeness of previously published *Salicornia* genome resources should be supported by concrete evidence, and we have revised the text accordingly.

With respect to the chromosome-level genome of *Salicornia ramosissima* generated by the UK Tree of Life project, the authors state in the original publication that “the genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline”. However, at the time of our revision (September 2025), no gene annotation or gene models have been released for this assembly. We have therefore clarified in the manuscript that this resource remains unannotated, rather than implying incomplete annotation.

For earlier *Salicornia* genomes described as draft assemblies, including *S. bigelovii* and diploid *S. europaea*, the original publications explicitly refer to these as draft genome sequences, which lack chromosome-scale scaffolding and comprehensive annotation. We now explicitly distinguish draft assemblies from chromosome-level reference genomes in the revised text to avoid ambiguity.

We also acknowledge and apologize for the misreferenced citations. The draft genome assemblies are correctly referenced as:

<https://doi.org/10.1038/s41467-024-48595-5>,

whereas the unannotated chromosome-level genome is described in:

<https://doi.org/10.12688/wellcomeopenres.21552.1>.

The reference list and in-text citations have been corrected accordingly.

Finally, we note that the recently published telomere-to-telomere assembly of tetraploid *Salicornia europaea* from China (Duan et al., 2025)[<https://doi.org/10.1093/plphys/kiaf565>] reports a genome size of ~984 Mb, whereas a tetraploid *S. depressa* genome from the United States reports a genome size of 1.38 Gb (Kimball-Rhines et al., 2025)[<https://doi.org/10.1093/jhered/esaf062>]. Although both studies designate their subgenomes as A and B, our comparative analyses indicate that these subgenomes are not homologous, which may confound researchers in further study and utilization of genomes. Based on comprehensive synteny and comparative analyses across our six assemblies, we identified four distinct subgenomes (A, B, C, and D), and found that the A and B subgenomes reported in *S. europaea* and *S. depressa* correspond to different evolutionary lineages rather than shared subgenomic identities.

We further examined chromosome numbering and orientation consistency across these published genomes. During this process, we identified substantial inconsistencies in chromosome numbering and orientation in the published *S. depressa* genome relative to *S. ramosissima* and tetraploid *S. europaea*, as well as relative to the subgenome framework established in our study. We have clearly documented these inconsistencies and implemented a standardized chromosome naming and

orientation scheme, as described in Supplementary Note S1. This correction provides a consistent comparative framework and facilitates accurate cross-study comparisons, improving the reliability and future utility of *Salicornia* genomic resources.

Together with unresolved subgenome structure and ongoing taxonomic ambiguity within *Salicornia*, this reinforces our statement that it remains unclear whether existing reference genomes are representative of local or global *Salicornia* diversity. We have revised the text to frame this as a limitation of current resources, rather than a criticism of prior studies.

The study lacks paired end shortread sequencing, which is essential for genome polishing and consequently fails to perform the necessary polishing to correct residual errors.

Response: We thank the reviewer for raising this point. All assemblies in this study were generated using PacBio HiFi (CCS) reads, which already provide very high per-base accuracy. According to the evaluation in Mascher et al. [<https://doi.org/10.1093/plcell/koab077>], polishing a CCS-derived assembly with Illumina short reads did not lead to measurable improvement in consensus accuracy and in fact slightly reduced the proportion of perfect gene model alignments. Similarly, the introduction of NextPolish2 emphasizes that conventional short-read polishing methods often fail to improve and may introduce overcorrections when applied to highly accurate HiFi/CCS assemblies [<https://doi.org/10.1093/gpbjnl/qzad009>].

To rigorously assess assembly accuracy, we quantified consensus quality values (QV) for all assembled chromosomes. Each chromosome exhibits $QV > 50$, corresponding to an estimated per-base error rate of $<1 \times 10^{-5}$, indicating a very low residual error rate and obviating the need for additional short-read polishing. These metrics are now explicitly reported in **Table 1**, the Results subsection “*Chromosome-scale genome assemblies and annotation quality*”, and **Supplementary Note 9**. We used Illumina reads to estimate polymorphism rates (% error) by mapping reads from each accession to its own assembly and observed a negligible error rate (Extended Data Fig. 3).

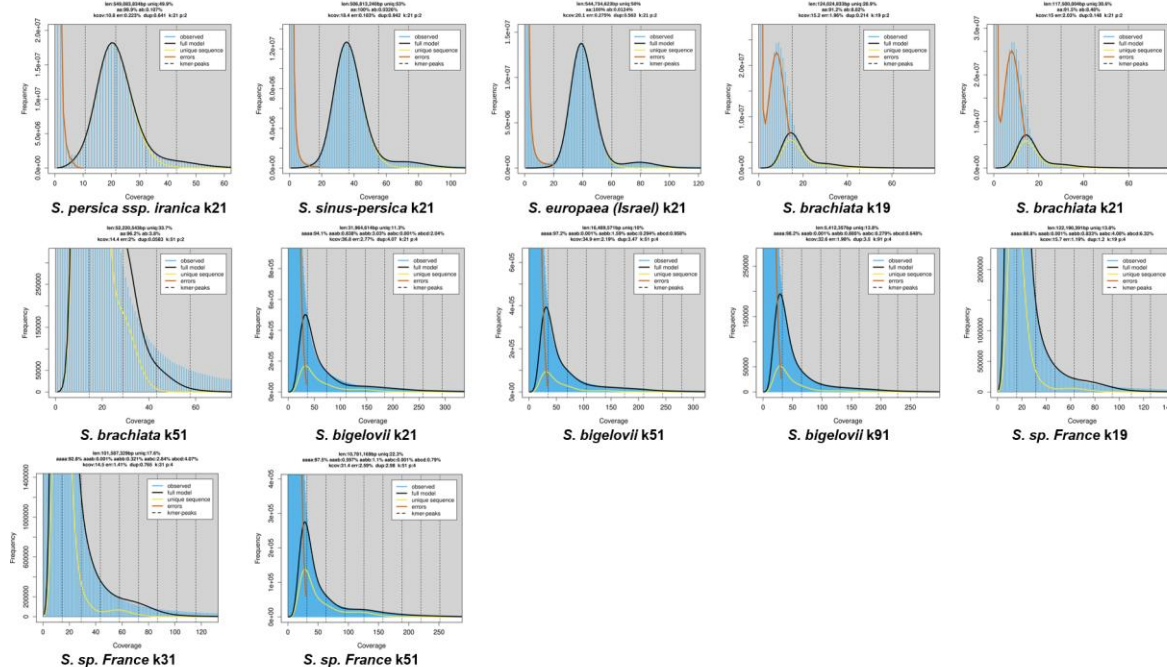
Given the multiple ploidy levels among the sampled species, pairedend shortread data are also indispensable for accurate genome size estimation and ploidy inference, thereby ensuring a more complete and robust body of evidence. [<https://www.nature.com/articles/s41467-020-14998-3>]

Response: In this study, we experimentally estimated genome size by measuring nuclear DNA content using flow cytometry with an internal standard (see **Supplementary Table 3** for pg DNA and reported genome size). We further employed diploid chromosome probes in hybridization experiments on tetraploid chromosome samples to inform ploidy inferences. Together these approaches provide independent and orthologal lines of evidence for genome size and ploidy determination.

Specifically, nuclear DNA content measurements yield p1C genome-size estimates that are highly informative for ploidy inference. While paired-end short reads and k-mer-based analyses can offer additional resolution—particularly for repeat structure and assembly guidance.

Despite extensive exploration of parameters, GenomeScope consistently produced unstable and implausible fits for our assembled *Salicornia* genome. In addition, we observed that GenomeScope

misclassified the known allotetraploid *Salicornia europaea* (China) as an autotetraploid (Duan et al., 2025). Similarly, for the published *Salicornia depressa* genome, GenomeScope estimated a genome size of ~282 Mb, which is markedly smaller than the final assembled genome size of 1.39 Gb (Kimball-Rhines et al., 2025). Given these inconsistencies, we considered the corresponding GenomeScope estimates to be unreliable and therefore did not include them in this manuscript.



The karyotype presentation in Extended Data Fig. 3 is unclear and lacks sufficient resolution.

Response: Thank you for comment. We have updated **Extended Data Fig. 3** by replacing it with higher-resolution karyotype images.

Line 139-141. The authors note that karyotyping of *S. europaea* (France) was unsuccessful. Please consider applying multiple techniques and repeating the assays to achieve a successful karyotype.

Response: Thank you for your suggestion. Following this advice, we applied additional cytogenetic analyses using fluorescence in situ hybridization (FISH). Specifically, we designed two arm-specific probes targeting the two different arms of Chr9C, together with a 45S rDNA probe located on Chr9C. These probes were hybridized to metaphase chromosome spreads of *Salicornia* sp. France (referred to as *S. europaea* (France) in the original manuscript).

The FISH results consistently revealed that *Salicornia* sp. France has $2n = 36$ chromosomes. The 45S rDNA signal is located in a central region of Chr9C and exhibits incomplete chromatin condensation and weak staining. We infer that this feature of the rDNA region likely caused the previously unsuccessful karyotyping using conventional staining alone. Representative FISH images and methodological details have been added to **Figure 2** and **Extended Data Fig. 4** and methodology to **Supplementary Note 10**.

Line 166-170. For subgenome homology analysis, it is standard practice to include appropriate

outgroup taxa and to build phylogenetic trees using aligned CDS data, rather than relying solely on OrthoFinder's fasttree mode.

Response: Thank you for suggestion. Following your recommendation, we revised our analysis by including one additional species from the Amaranthaceae family, *Spinacia tetrandra*, and one species outside the family, *Arabidopsis thaliana*, as outgroup taxa. We identified 4,977 single-copy orthologous genes across the genomes of *Spinacia tetrandra*, *Arabidopsis thaliana*, *S. brachiata*, *S. sinus-persica*, *S. europaea* (Israel), *S. persica ssp. iranica*, *S. sp.* (France) subgenomes A and C, and *S. bigelovii* subgenomes B and D (data shown in **Fig. 2b**). Multiple sequence alignments (MSA) were performed using the coding sequences (CDS) of these single-copy orthologs, and species trees were subsequently constructed based on the concatenated MSA results. The corresponding contents are added in "Materials and methods" section titled "Subgenomes phylogenomics comparison", as well as in Results under sub-heading "*Karyotype validation and subgenome resolution reveal tetraploid genome organisation in Salicornia*" and Discussion under sub-heading "*Polyploid origin and subgenome evolution in Salicornia*".

In Fig. 3b, why do intraspecies paralogous genes in the diploid taxa exhibit a recent Ks peak, particularly in *S. brachiata*?

Response: After inspecting the paralog pairs that contribute to the recent Ks peak in *S. brachiata*, we found that the signal is predominantly explained by tandemly duplicated gene pairs rather than by genome-wide syntenic blocks. Tandem duplicates often produce pronounced, relatively young Ks peaks and can therefore be mistaken for signals of whole-genome duplication. To illustrate, lineage-specific tandem expansions in sunflower have been documented and dated using Ks distributions, showing similar recent peaks that were attributed to tandem (small-scale) duplications (Badouin et al. 2017 <https://doi.org/10.1038/nature22380>). Accordingly, we interpret the small recent peak in *S. brachiata* as arising from tandem duplication. Importantly, a similar near-zero Ks peak is present in *S. sinus-persica*, *S. europaea* (Israel) and *S. persica ssp. iranica*; in **Fig. 3b** these signals were visually obscured by the *S. brachiata* trace due to color overlap. To clarify this point we have added species-separated Ks plots and an annotated breakdown of paralog pairs (tandem versus syntenic) in the Supplementary **Supplementary Fig. 4**, which demonstrate that the recent peaks in all four taxa are driven primarily by lineage-specific tandem (small-scale) duplications.

In Fig. 3a, although Line 205-206 describes "the divergence of the diploid progenitors", the figure provides no annotation to clarify this, which may lead readers to misinterpret the two circles as marking the timing of polyploidization events. Moreover, directly plotting progenitor divergence times on the phylogenetic tree is unconventional and should be reconsidered.

Response: Thank you for the helpful comment. We agree that the original annotation could be misunderstood and have revised the figure and accompanying text to avoid this confusion. Specifically, we have modified **Fig. 3a** so that polyploid taxa are shown by their resolved subgenomes rather than using circles that might be interpreted as timing of polyploidization. We also removed the direct plotting of inferred progenitor-divergence times on the main tree.

Line 208-212. The results in Fig. 2c and Fig. 3c are inconsistent: Fig. 2c reveals a larger number of chromosomal structural rearrangements, whereas they should display the same overall pattern.

Response: The apparent discrepancy arises because **Fig. 2c** and **Fig. 3c** visualize genome organization at different scales and with different unit definitions. Fig. 2c presents structural variation at the genomic DNA level, including fine-scale rearrangements, whereas Fig. 3c displays conserved syntenic blocks defined as runs of 20 consecutive homologous proteins with preserved order across species. As a result, small structural rearrangements that are evident in Fig. 2c do not form independent blocks and are therefore not shown in Fig. 3c; only large-scale rearrangements affecting block order are visible. Importantly, the primary purpose of Fig. 3c is to reconstruct ancestral karyotype structure rather than to quantify genome-wide structural variation. We have added this clarification to Supplementary Note 15.

The population genetics analysis is relatively simple and conventional, maybe constrained by low sequencing depth. They fail to resolve the *Salicornioideae* subfamily taxonomy or to clarify the origin and evolutionary history of the genus.

Response: We thank the reviewer for this important comment. The population genetic analyses were designed to provide a robust assessment of phylogenetic relationships, population structure, and genome-wide diversity across a globally representative *Salicornia* panel, and we have included additional analyses in the revised manuscript to further strengthen these conclusions. The sequencing depth (mean $\sim 10\times$ per accession) is well within the accepted range for population genomic studies and enabled the identification of millions of high-confidence SNPs, providing sufficient resolution to characterize genetic diversity and genomic relationships among accessions.

However, the panel has certain limitations. It includes multiple species, broad geographic representation, and polyploid lineages with distinct subgenome compositions, resulting in strong population structure across ploidy levels and reflecting the inherent complexity of the genus. Therefore, our study was not designed to resolve the *Salicornioideae* subfamily taxonomy or fully clarify the evolutionary history of the genus; rather, it focuses on generating chromosome-scale reference genomes and characterizing global genomic diversity using SNP- and k-mer-based analysis.

Line 323. The gene function analysis lacks innovation, relying solely on basic transcriptome profiling and the selection of a handful of differentially expressed genes. By contrast, the earlier Nature Communications paper, titled “SOS1 tonoplast neo-localization and the RGG protein SALTY are important in the extreme salinity tolerance of *Salicornia bigelovii*” offers far deeper mechanistic insights and experimental logic.

Response: We thank the reviewer for this comment. We have removed this analysis because the evidence is insufficient to support robust conclusions, and the reliability of results inferred from publicly available datasets cannot be fully guaranteed.

Line 337-344 FPKM is a relative measure of gene expression, and its application may lead to inaccurate results when used to compare expression levels across different species.

Response: We thank the reviewer for this valid point. We have removed the analyses and results that relied on cross-species FPKM comparisons; consequently, there are no longer any relative-expression (FPKM) claims in the manuscript.

Point-by-point responses to reviewers' comments

Please note that line numbers refer to the manuscript with track-changes provided.

Reviewer #1 (Remarks to the Author):

The revised manuscript is further improved relative to the first submission, moving the work beyond a “resource announcement” toward a more hypothesis-driven evolutionary genomics study. However, several issues still limit interpretability and should be addressed before the manuscript can be accepted.

Major concerns

1. Selection inference at the OSCA locus requires additional support.

The OSCA region is presented as a central adaptive result, with contrasting Tajima's D between *Salicornia* sp. France and *S. procumbens* interpreted as evidence for a selective sweep versus balancing selection. However, Tajima's D is sensitive to demography, population structure, uneven sampling, and genotyping/mapping bias, and on its own is insufficient to conclude balancing selection, particularly in complex polyploid-derived systems. Please provide complementary evidence in the OSCA region, such as local nucleotide diversity (π), a sliding-window F_{ST} scan between the two tetraploid groups, or a composite-likelihood (CLR) type statistic. In addition, please present allele-frequency patterns and local haplotype structure around OSCA to demonstrate that the signal is not driven by local mapping bias or missing data.

Response: We agree that Tajima's D alone can be influenced by demographic shifts, variable sample sizes, or sequence mapping artifacts. To address this concern comprehensively, we have extended our analyses at the OSCA locus with four complementary lines of evidence: (1) empirical down-sampling controls to rule out sample-size bias; (2) nucleotide diversity (π) profiling that independently corroborates the contrasting signals between species; (3) cross-validation against a publicly available *S. ramosissima* reference genome to eliminate mapping and assembly artefacts; and (4) high-resolution haplotype-block visualisation confirming that the divergent population-genetic signatures reflect genuine evolutionary history rather than data quality issues. Taken together, these analyses provide robust, multi-method support for our interpretation of a directional selective sweep in *S. procumbens* and balancing selection in *Salicornia* sp. France at this locus.

1) Addressing sample size bias (down-sampling control):

To empirically test whether uneven sampling sizes between the two species skewed our statistics, we randomly down-sampled the *Salicornia* sp. France population ($N = 71$) into independent subsets matching the sample size of *S. procumbens* ($N = 12$). Re-running the Tajima's D scans across these size-matched cohorts reproduced identical, contrasting distribution patterns on chromosome 6A. This confirms that the observed divergence is not an artifact of sample-size imbalances or uneven sampling density (now detailed in **Lines**

472-478). Tajima's D values derived from 12 randomly sampled accessions of *Salicornia* sp. France and *S. procumbens* are shown below (Figure 1).

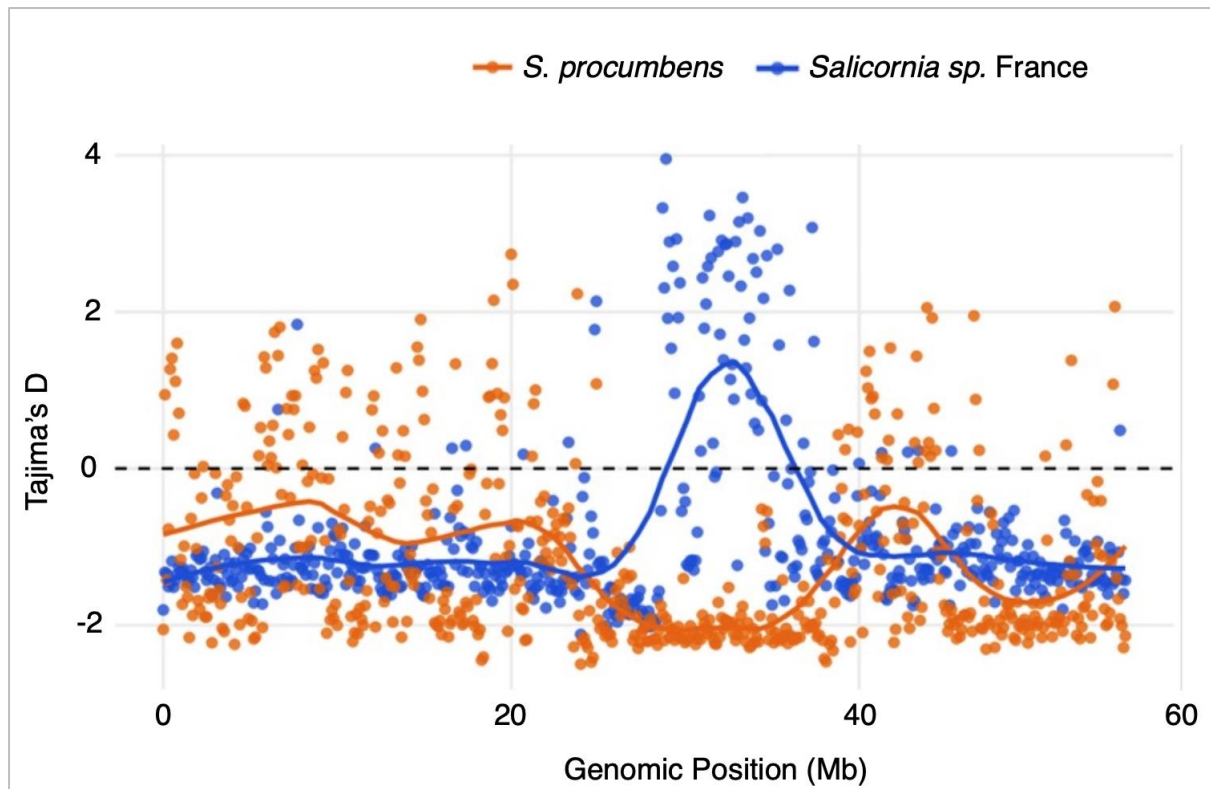


Figure 1. Tajima's D values derived from 12 random samples of *Salicornia* sp. France and *S. procumbens*. Blue and orange dots represent window-based estimates for *Salicornia* sp. France and *S. procumbens*, respectively, with solid lines indicating LOESS-smoothed trends.

2) Complementary population-genetic evidence (nucleotide diversity, π):

We computed nucleotide-diversity (π) profiles. Local π scans revealed *Salicornia* sp. France has a sharp elevation in localized π , which aligns with its highly positive Tajima's D peak, whereas *S. procumbens* displays a severe depression in local diversity tracking its negative Tajima's D plunge (**Extended data Figure 10a, supplementary Fig. 17**).

3) Ruling out reference and mapping bias via cross-assembly validation:

To definitively eliminate the possibility that the observed selection signal was an artifact of structural collapses, reference deletions, or multi-copy cross-mapping within the draft *Salicornia* sp. France reference assembly, we re-mapped all WGS short-read and re-called variants against an independent, highly contiguous, diploid *S. ramosissima* reference genome. Strikingly, the mirror-image Tajima's D profiles and the localized π contrast were recapitulated on the independent reference background (**Extended data Fig. 10a and Supplementary Fig S17**). This cross-genome validation rules out reference-specific assembly topology artifacts.

4) High-resolution local haplotype structure and allele-frequency patterns:

To demonstrate that the signatures ~33–35 Mb on chromosome 6A are driven by genuine evolutionary haplotypes rather than missing data or mapping noise, we generated a haplotype-block visualization at sample levels and annotated the true genomic coordinates of the OSCA exons around 35.8 Mb of *S. ramosissima* genome (**Extended data Figure 10b**). *Salicornia procumbens* accessions display an almost completely uniform, fixed haplotype block across the target region, most likely validating our inference of a directional selective sweep under a predictable environmental regime. On the other hand, *Salicornia* sp. France accessions segregating with either ref, alt or het alleles and localized structural-deletion variants (missing tracks). This mosaic-type haplotype blocks demonstrates that balancing selection in the fluctuating French marsh environment is actively preserving diverse structural haplotypes to maintain intermediate allele frequencies that drive the localized spike in both Tajima's D and π (**Extended data Figure 10b and Supplementary Figure 15**)

2.SNP calling strategy for the tetraploid A-subgenome should be described more transparently. The manuscript states that SNP-based analyses were conducted on the A-subgenome of the tetraploids to compute π and Tajima's D, but the calling procedures and filtering criteria are not sufficiently detailed in the main text. Please clarify: (i) which reference was used (i.e., the source assembly and specific A-subgenome chromosome set), (ii) how homeologous mapping was prevented or minimized (e.g., unique-mapping thresholds, masking strategy, mapping-quality filters), (iii) whether genotypes were treated as diploid when analyzing the A-subgenome and the justification for this assumption, (iv) the full filtering criteria applied to derive the tetraploid A-subgenome SNP set (e.g., QUAL/DP/MQ thresholds, missingness, MAF), and (v) the sample size (n) used for π and Tajima's D calculations (ideally reported in the relevant figure legends).

Response: We have expanded and clarified the subgenome-specific SNP-calling, mapping, and filtering strategies in the revised Methods section (**Lines 969-974 and Supplementary Note 17**). In summary:

For nucleotide diversity (π) and Tajima's D analyses, raw whole-genome sequencing (WGS) reads from tetraploid AC-lineage accessions (*Salicornia* sp. France and *S. procumbens*) were mapped exclusively to the chromosome-scale *Salicornia* sp. France reference assembly. We now include the analysis using the *S. ramosissima* genome to map A-subgenome-specific reads of *Salicornia* sp. France and *S. procumbens* (**Extended Data Fig 10**)

To minimize ambiguous homoeologous alignment and eliminate non-specific read mapping between subgenomes, we implemented a stringent bioinformatic filtering regime. Sequence reads were aligned using HISAT2, and the resulting alignments were processed using SAMtools to retain strictly unique and concordantly mapped read pairs. Multi-mapped, non-specific, and improperly paired read tracks were excluded prior to and during variant-calling procedures.

Empirical mapping statistics [**Supplementary Figure 14 and Extended Data Fig. 8b** (normalized read-depth graph)] verified this unique-mapping protocol. We cross-examined the alignment specificity of representative AC (*Salicornia* sp. France; SalD_100) and BD (*S.*

bigelovii; SalD_109) subgenome accessions against a complete *in silico* octaploid assembly. Mapping statistics demonstrate extreme specificity (> 98%) to the relevant chromosomes. This minimal cross-subgenome mapping ($\leq 2\%$) confirms that homeologous noise is neutralized at the target locus.

Genotypes within the isolated subgenome tracks were treated as diploid during variant calling. This approach is biologically and analytically justified because the subgenomes represent discrete, historically independent chromosome sets within the allotetraploid framework.

Raw variants were filtered through a rigorous, multi-step pipeline using 'bcftools' and custom in-house scripts to derive a high-confidence biallelic SNP set. Only variants passing all of the following sequential filters were retained for downstream population genetics calculations:

- Quality and global depth: Initial filtering using 'bcftools filter' retained only sites satisfying a quality score of $QUAL \geq 30$ and an overall read depth of $INFO/DP \geq 20$.
- Allele frequency: Sites were filtered to retain variants with a baseline allele frequency range of $0.01 < AF < 0.99$ and a Minor Allele Frequency threshold of $MAF > 0.01$.
- Data integrity and heterozygosity: Variants were strictly filtered to exclude sites with missing data exceeding 20% (< 20% missingness) or individual loci displaying an inflated heterozygosity fraction (> 20% heterozygosity) to eliminate potential collapsed paralog/homeolog mapping artifacts.

Minor concerns

1. Please clarify and justify the outgroup choice used in each set of tests and ensure consistency across analyses.

Response: In the revised manuscript, we have clarified the rationale for outgroup selection and ensured a consistent strategy across all phylogenetic analyses in **Supplementary Note 13**.

For the family-level analysis of Amaranthaceae (**Figure 2B**), *Arabidopsis thaliana* was selected as the outgroup because (i) as a dicotyledonous species, it provides an appropriate evolutionary distance for reliably rooting the ingroup; and (ii) it possesses the most comprehensive and highest-quality genome annotation currently available, ensuring accurate identification of orthologous sequences.

For the genus-level analysis within *Salicornia* (**Figure 3A**), we adopted a two-tiered outgroup strategy. *Arabidopsis thaliana* was retained as the distant outgroup to maintain consistency with the family-level analysis, while *Spinacia oleracea* was additionally included as a closely related outgroup. *Spinacia tetrandra* was chosen because it represents the closest available telomere-to-telomere (T2T) genome assembly within Amaranthaceae outside of the genus *Salicornia*, thereby providing a robust and well-annotated near-outgroup for resolving fine-scale relationships within the genus.

This hierarchical outgroup scheme — using *Arabidopsis thaliana* as a consistent distant outgroup across all analyses, supplemented by a near-outgroup at the genus level wherein

finer resolution is required — ensures that tree topologies are directly comparable while maximizing phylogenetic accuracy at each taxonomic scale.

2. Please soften deterministic phrasing such as “balancing selection” and “rapid fixation of an optimal OSCA allele,” as these are strong conclusions without demography-aware testing.

Response: We have softened the deterministic phrasing throughout the manuscript to better reflect the inferential nature of these population-genetic signatures. We have revised "balancing selection" to "a signature consistent with balancing selection" and revised "rapid fixation of an optimal OSCA allele" to "a selective sweep consistent with a locally advantageous OSCA allele."

Further discussion is provided on **Lines 574-582**.

3. The revised manuscript still contains minor typographical issues (e.g., duplicated words such as “Notably, Notably,” spelling inconsistencies like “S. europea” vs “S. europaea,” and minor spacing/formatting problems). A careful final proofread is recommended.

Response: These have been addressed.

4. Please also provide the CDS and protein fasta files of each genome assembly in the DRYAD.

Response: We thank the reviewer for this important reminder. We have updated CDS and protein fasta files of each genome assembly in the DRYAD.

Comments on the authors’ response to Reviewer #2

The authors have made a relatively serious effort in responding to Reviewer #2. However, there remain two points that require further attention and clarification.

(1) On the response to the reviewer’s comments regarding the Introduction
In the Response, the authors emphasize the question of whether genomic changes such as structural variation, copy-number variation, and gene-family remodeling are associated with adaptation to extreme saline environments. In the revised manuscript, this is reflected in the aim: "(iv) to identify candidate genes associated with salinity-related traits through comparative genomics within Salicornia and across the Amaranthaceae".

However, in practice, the results appear rather limited. The selection analyses seem to identify only a small number of candidates (e.g., OSCA gene family), followed by a brief discussion linking these to soil electrical conductivity (which presumably reflects salinity). As a result, this part of the manuscript has become somewhat underdeveloped and lacks depth.

Response: To address the concern that gene-family remodelling has not been explicitly linked to salinity adaptation, we have substantially extended the comparative analysis. We now present evidence across three gene families: the OSCA family, which shows the strongest signal of divergent selection between *Salicornia* species contrasting in salinity tolerance; and two families showing the strongest lineage-specific expansion in our orthology dataset; chalcone synthase (*CHS*, expanded in the *Salicornia* genus) and *Cupin 1* (expanded across *Salicornia*, *Suaeda*, and *Haloxylon*).

For the OSCA family, our population-genomic analyses identified a striking divergence signal on chromosome 6A, with contrasting Tajima's D and π profiles between *Salicornia* sp. France and *S. procumbens* (**Extended Data Fig 10a**), consistent with divergent selective regimes acting on this locus. This signal, supported by haplotype-block analysis (**Extended Data Fig 10b**) and robust to down-sampling controls (see response to **Reviewer #1** major comment 1.), points to the OSCA family as a strong candidate underlying differences in salinity-related traits between these lineages, further corroborated by the association with water electrical conductivity measurements.

For the *CHS* and *Cupin 1* families, we asked whether the expanded copies show *Salicornia*-specific transcriptional responses under salt stress that distinguish them from other Amaranthaceae genera halophytes. We re-analyzed our lab's published RNA-seq data from *S. bigelovii* shoots harvested after 6 weeks of 600 mM NaCl treatment (Salazar et al., *Nat. Commun.* 2024) and publicly available data from *Amaranthus hypochondriacus* leaves harvested at the end of a 14-d stepped salinity regime (0.15 M $\times$ 7 d $\rightarrow$ 0.3 M $\times$ 4 d $\rightarrow$ 0.5 M $\times$ 3 d; Barba de la Rosa et al., *Planta* 2026). Both tissues accumulate Na⁺ under salt stress, since the treatment durations (6 weeks and 14 d, respectively) are long enough for systemic Na⁺ translocation and tissue-level (i.e. shoot) accumulation. The two species belong to the same family (Amaranthaceae) but contrast sharply in Na⁺-accumulation capacity — *S. bigelovii* being a strong Na⁺ accumulator and *A. hypochondriacus* a comparatively weak one (*S. bigelovii* shoots accumulate Na⁺ to >750 mM at 600 mM NaCl, whereas leaves of grain amaranths reach only ~100–150 mmol kg⁻¹ DW under salt stress, roughly two orders of magnitude lower; Estrada et al., 2021) — enabling a within-family cross-species comparison of how the lineage-expanded *CHS* and *Cupin 1* families respond in salt-stressed, Na⁺-accumulating tissues. Orthologs of each *S. europaea* *CHS* and *Cupin 1* anchor were identified by reciprocal BLASTp ($E < 1e-30$; $\geq 60\%$ identity with $\geq 70\%$ query coverage, or $\geq 80\%$ identity with $\geq 30\%$ query coverage). DEGs were called with DESeq2 (Wald test, BH-adjusted $P < 0.05$, log2FC magnitude > 0.585) independently within each species against its own controls. The new analyses are shown in **Supplementary Figs. 7b, c** and discussed on **lines 309–320**.

- *CHS* family (**Supplementary Fig. 7b**): Despite the lineage-specific expansion, no *S. bigelovii* *CHS* ortholog reaches DEG statistical significance ($P_{adj} < 0.05$) in either the B or D subgenome, across both timepoints and salt concentrations. In contrast, both *A. hypochondriacus* *CHS* orthologs are statistically significant DEGs ($P_{adj} < 0.001$).

This indicates that the expanded *Salicornia* CHS family is not transcriptionally responsive to salt at the DEG level, unlike the *Amaranthus* orthologs.

- **Cupin 1 family (Supplementary Fig. 7c):** Among the *S. bigelovii* Cupin 1 orthologs, only one gene — SaBi02252 in the B subgenome — reaches DEG statistical significance ($P_{adj} < 0.001$); no ortholog in the D subgenome passes the threshold. *Amaranthus hypochondriacus* shows no statistically significant DEG in this family. The single statistically significant DEG, restricted to the B subgenome of *Salicornia*, points to a narrow, B-subgenome-biased transcriptional response on the lineage-expanded copies rather than a family-wide DEG signal.

Together, these results demonstrate that candidate genes and gene families linked to salinity adaptation in *Salicornia* have undergone changes at multiple levels — divergent selection (OSCA), copy number expansion (CHS, Cupin 1), and salt-responsive expression (Cupin 1, B subgenome) — a multi-layered pattern that differs markedly from the non-accumulating *A. hypochondriacus* and substantiates the depth of the comparative-genomic analysis. We trust these additions now add adequate depth to this section.

(2) On the response to the reviewer's comments regarding transcriptome analysis and species selection

Reviewer #2 raised several important points regarding the intra-family transcriptome comparison ("Lines 417-429: All the effort of intra-family transcriptome comparison ..."), including the question "Why did you select *S. bigelovii* for more detailed gene expression analysis?". In the current Response, the authors emphasize practical limitations. However, the key question regarding the rationale for selecting *S. bigelovii* for detailed expression analysis has not been directly addressed.

Response:

Why *S. bigelovii* was selected: Among the six *Salicornia* genomes assembled in this study, *S. bigelovii* was the only species for which a published, controlled, salt-treatment RNA-seq dataset existed in the public domain at the time of submission (Salazar et al., 2024). No dedicated salt-stress RNA-seq experiment was performed as part of the present study.

Why the comparison was removed: Upon re-evaluation, we concluded that the three public datasets — while nominally generated under "salt treatment" — were designed to address fundamentally different biological questions, capturing early salt-signalling responses (Yu et al., 2022; 48 h, 50 mM NaCl), peak active stress responses (Hosseini et al., 2023, 4 d, 13.8 dS.m⁻¹), and long-term salt-acclimation homeostasis (Salazar et al., 2024; 6 weeks of 200mM NaCl) respectively. Directly comparing ortholog expression across these datasets conflates experimental objective with interspecific contrast, meaning any apparent *Salicornia*-specific signature could reflect experimental-design differences rather than genuine lineage-specific regulation. Removal was therefore the only defensible course of action.

Directly comparing ortholog expression across these three datasets therefore conflates three different experimental objectives (early salt-response signalling vs. peak active stress response vs. long-term salt-acclimation homeostasis) with the interspecific contrast we intended to test (high vs. low salt accumulator). Any apparent "*Salicornia*-specific" expression signature emerging from such a comparison could be confounded by this

mismatch in experimental purpose rather than by genuine lineage- or trait-specific regulation.

Our new comparative transcriptome analysis (*CHS*, *Cupin 1*; **Supplementary Figs. 7b,c**) does not share these vulnerabilities because both datasets used in that comparison capture tissue-level Na⁺ accumulation under sustained salt exposure.

Reviewer #3 (Remarks to the Author)

The authors have made revisions, and I appreciate the effort they have put into addressing the reviewers' comments. However, some of the changes appear overly expedient and, in my view, have even compromised the overall quality and integrity of the manuscript. In particular, several reviewers raised clear concerns regarding the RNA-seq analysis of the *SOS1/2* genes, noting significant methodological issues. Instead of properly addressing these concerns, the authors have chosen to remove the relevant results entirely. This approach seems overly simplistic and somewhat evasive. Moreover, as the authors emphasize in both the title and throughout the manuscript, the halophytic characteristics of the genus *Salicornia*, the removal of this key functional analysis, substantially weakens the central biological narrative. Replacing it with a relatively simple comparison of gene family copy numbers is insufficient to support the claims made, particularly for a study aiming to meet the standards of *Nature Communications*.

Response: We respectfully disagree with the suggestion that the revisions were “overly expedient”, “simplistic”, or “evasive”. In response to the helpful and detailed reviewers' comments, we undertook extensive additional analyses and substantial restructuring of the manuscript to improve both its rigor and biological interpretation. These revisions included:

- (i) re-orientation, renumbering, and harmonization of previously published and newly generated *Salicornia* genomes to establish a standardized subgenome framework;
- (ii) expanded structural-variation and copy-number analyses across *Salicornia* and related *Amaranthaceae* lineages;
- (iii) additional population-genomic analyses including *FST* and Tajima's *D* and haplotype-block analysis;
- (iv) codon-aware phylogenetic re-analysis of subgenome relationships; and
- (v) new transcriptomic analyses of lineage-expanded *CHS* and *Cupin 1* families under biologically comparable long-term salt-acclimation conditions. Collectively, these revisions substantially extend the manuscript beyond a genome-resource study and integrate comparative genomics, population genomics, transcriptomics, and environmental metadata to investigate the evolutionary basis of halophyte adaptation in *Salicornia*.

We also wish to clarify that removal of the previous cross-species *SOS1/2* expression analysis was not undertaken to avoid criticism, but rather because we concluded that the analysis could not be interpreted with sufficient rigor given the substantial differences in experimental design among the public RNA-seq datasets used in that analysis. Those datasets represent fundamentally different biological states, including early salt-signalling responses, peak stress-response phases, and long-term acclimation homeostasis. Under such conditions, apparent interspecific expression differences could not be confidently separated from differences arising from experimental design. For this reason, the *SOS1/2* expression analysis was removed from the revised manuscript. We considered removal of this analysis to be the most scientifically rigorous and transparent course of action.

Importantly, rather than simply removing this section, we replaced it with new analyses based on biologically comparable long-term salt-acclimation datasets. To further strengthen the biological interpretation of the lineage-specific expansion of *CHS* and *Cupin 1* gene families in *Salicornia*, we identified a suitable public transcriptomic dataset from Barba de la Rosa et al. (2026) examining long-term salt acclimation in the salt-sensitive Amaranthaceae species *Amaranthus hypochondriacus*. In that study, plants were maintained for one month under control conditions before gradual NaCl increases over 11 d to a final concentration of 0.5 M NaCl, followed by continued exposure for 3 d prior to leaf sampling. This experimental design captures sustained acclimation-phase transcriptional responses rather than transient early salt-signalling events. This design is biologically comparable to our previously published *S. bigelovii* transcriptomic dataset (Salazar et al. 2024) generated under long-term salt-treatment conditions in the same tissue type, while also preserving a biologically informative contrast between a low-Na-accumulating salt-sensitive Amaranthaceae species (*A. hypochondriacus*) and the extreme halophyte *Salicornia*.

Using reciprocal orthology assignment and differential-expression analysis under matched acclimation conditions, we found that the lineage-expanded *CHS* and *Cupin 1* paralogs in *Salicornia* exhibit coordinated transcriptional remodelling during sustained salt exposure. These expression patterns differed substantially from those observed in *A. hypochondriacus*, indicating that the expanded paralogs are not simply products of neutral genome expansion but represent biologically responsive gene-family components associated with long-term halophytic acclimation. These analyses are now presented in **Supplementary Fig. 7b and c** and discussed in detail in the response to **Reviewer #2**, Question 1.

While comprehensive functional validation of candidate loci is beyond the scope of the present comparative-genomics study, we agree that connecting genome evolution to biologically meaningful adaptive hypotheses is important. Our aim in this work was therefore instead to first establish a robust evolutionary and phylogenetically informed genomic framework from which candidate adaptive loci, pathways, and subgenomic innovations can now be rigorously tested in future hypothesis-led functional studies.

More comments are listed below.

For the subgenome homology analysis, although the authors state that they used CDS data, the CDS sequences appear to have been directly aligned and used for phylogenetic inference. This approach is not sufficiently rigorous. Direct alignment of CDS can introduce alignment errors due to frame shifts and codon structure. A more reliable strategy is to first align the corresponding protein sequences of single-copy orthologs, and then translate the protein alignment to CDS alignments using PAL2NAL. After that, ambiguously aligned regions should be filtered out before tree construction. This workflow ensures codon-aware alignment quality and reduces phylogenetic artifacts. All phylogenetic analyses in the manuscript should follow this more rigorous strategy to ensure robustness and reproducibility of the inferred evolutionary relationships.

Response: Thank you for this recommendation. We have reanalysed the subgenome homology dataset using this pipeline (**Supplementary Fig. 22**).

We re-aligned all 4,977 single-copy orthologous groups at the protein level using the MAFFT alignments produced by OrthoFinder, and back-translated these protein alignments to codon-aware CDS alignments using PAL2NAL v14. Ambiguously aligned regions were then filtered with trimAl v1.5 (-automated1), with the trimming projected onto the codon coordinates via the -backtrans option (-ignorestopcodon to handle terminal stop codons) so that codon integrity is strictly preserved. After excluding seven orthogroups for the CDS of which could not be unambiguously back-translated, the remaining 4,970 codon-aware alignments were concatenated into a supermatrix of 7,442,829 bp across 10 taxa. Phylogenetic inference was performed with IQ-TREE 2 under the best-fit model selected by ModelFinder (GTR+F+R10, BIC), with branch support assessed by 1,000 ultrafast bootstrap replicates and 1,000 SH-aLRT replicates. The resulting phylogeny is topologically identical to the tree previously reported in the manuscript, and is also fully congruent with the single-copy protein-based phylogeny we presented in the first version of this study. All internal branches receive maximal support (UFBoot = 100, SH-aLRT = 100). The convergence of these three independent analyses — direct CDS alignment, protein-based inference, and codon-aware PAL2NAL/trimAl reconstruction — provides strong and consistent evidence for the inferred evolutionary relationships, and we believe this fully addresses the reviewer's concern regarding alignment rigour and reproducibility.

The authors attribute the recent Ks peak of intraspecies paralogous genes in the diploid taxa, particularly in *S. brachiata*, to tandem duplication events and provide Supplementary Fig. 4 as supporting evidence. However, this explanation is not clearly described or discussed in the main manuscript. Given that the presence of a recent Ks peak in diploid taxa could easily lead readers to speculate about additional whole-genome duplication events or other large-scale processes, it is important that the authors explicitly clarify this point. I recommend that the explanation regarding tandem duplication be incorporated into the relevant Results section or discussed more thoroughly in the Discussion to avoid confusion.

Response: We have now incorporated an explicit explanation into the Results section "Polyploid evolution and repeat-driven genome expansion in *Salicornia*" (Lines 243-248), clarifying that this recent Ks peak reflects lineage-specific tandem duplication events rather than additional polyploidization. The added text reads as follows:

"Notably, all diploid species also displayed a recent Ks peak among intraspecies paralogous gene pairs (Supplementary Fig. 4). However, these recently duplicated paralogs were predominantly organized as tandem arrays along chromosomes rather than dispersed genome-wide, indicating that this signal reflects lineage-specific tandem duplication events rather than an additional whole-genome duplication (Supplementary Fig. 4)."

We believe this revision directly addresses the reviewer's concern and improves the clarity of the interpretation.

In Supplementary Note 1, the dot plots shown in Figures 1 and 2 display a lot of background noise. Please check it and use a reliable cut-off.

Response: We have re-filtered the data using a strict minimum identity cut-off of 0.5 (50%) and removed the remaining low-identity background noise. The refined, higher-contrast dot plots have been uploaded to Figures 1 and 2 in Supplementary Note 1.

There appears to be an inconsistency in the use of italics for taxonomic names in both the text and figures. In several places, *Salicornia* and some Latin species names are not italicized (e.g., the second paragraph of Supplementary Note 1). Please carefully read the manuscript and all supplementary materials to ensure consistent formatting of genus and species names. In addition, for names such as *S. persica* ssp. *iranica*, please note that “ssp.” should not be italicized, whereas the genus/species epithet and the subspecies epithet should be italicized.

Response: The taxonomic names have now been italicised, and the subspecies abbreviation has been corrected to non-italicised font. These corrections were made throughout the manuscript.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

1. The authors have made a serious effort to address my previous concern about the OSCA locus, and the revised analyses are useful. The down-sampling control helps reduce concern about unequal sample sizes, the local nucleotide diversity profiles provide an additional layer of evidence beyond Tajima's D , and the remapping to an independent *S. ramosissima* reference is a helpful attempt to rule out reference-specific artefacts. The local haplotype-block visualization is also informative and strengthens the case that the OSCA region shows an unusual population-genetic pattern.

However, I do not think this issue has been fully resolved. My original concern was not only that Tajima's D required support from additional summary statistics, but also that the interpretation of balancing selection should be treated cautiously in a polyploid-derived system where demography, population structure, local structural variation, missingness, and residual homeologous mapping can all influence allele-frequency-based statistics.

The new analyses establish OSCA as a more credible candidate region, but they do not yet demonstrate balancing selection. In particular, I do not see the requested sliding-window F_{ST} analysis or a CLR-type selection scan across the OSCA region. These analyses would be important because the manuscript uses this locus as a central adaptive example. Without an independent statistic for local differentiation or selection, the current evidence is better described as a contrasting population-genetic signal or a selection-like signature, rather than as direct evidence that balancing selection is actively maintaining OSCA haplotypes.

I therefore recommend that the authors either add further quantitative support, such as a local F_{ST} scan and ideally a CLR/XP-CLR/RAiSD/SweeD-type statistic, or substantially soften the interpretation. A more appropriate conclusion would be that the OSCA region shows signatures consistent with divergent selection, or that it represents a candidate locus requiring further validation, rather than presenting it as a demonstrated case of balancing selection.

Response 1: We agree with the reviewer's assessment and have opted to substantially soften our interpretation rather than introducing additional selection scans that may be difficult to interpret given our modest population sizes (for one species we have only 12), polyploid genomic background, and the potential influence of demographic history and structural variation.

Specifically, we no longer present the OSCA region as evidence of active balancing selection or heterozygote advantage. Instead, we consistently describe it as a candidate adaptive genomic interval exhibiting a distinctive population-genetic signature that requires further validation.

To strengthen the descriptive evidence while avoiding over-interpretation, we additionally performed a sliding-window F_{ST} analysis across the chromosome 6A encompassing the OSCA region using balanced population sizes (**Supplementary Fig. 20**; 12 individuals per population). The analysis revealed localized variation in population differentiation across the interval, which

complements the contrasting Tajima's D and nucleotide diversity patterns without being interpreted as evidence for any specific selective process. These softened interpretations have been incorporated into the edited Results (L456-463; L483-491) and Abstract (L57-61).

We also substantially revised the Discussion (L528-585; L608-647). Rather than arguing for balancing selection, we now explicitly acknowledge that the observed population-genetic patterns may arise through multiple, non-mutually exclusive processes, including balancing selection, directional selection, demographic history, ancestral polymorphism, historical introgression, and local structural variation. Furthermore, we emphasize that although the genomic interval contains an OSCA/CSC1-like gene with a known role in osmotic stress signaling, the present data do not distinguish among these alternative mechanisms.

Finally, we revised the concluding discussion to state that resolving the evolutionary basis of the chromosome 6A region will require larger population samples, high-depth targeted or long-read phased sequencing, and pangenome-based analyses to accurately characterize local structural variation and haplotype architecture.

We believe these revisions provide a more balanced and appropriately cautious interpretation that aligns with the reviewer's recommendations while preserving the significance of the chromosome 6A OSCA region as a candidate target for future investigation.

2. The clarification of the A-subgenome SNP-calling strategy is also an improvement, but several technical details should still be made more explicit. The authors should report the mapping-quality threshold, per-sample depth filters, missingness and heterozygosity summaries specifically for the OSCA region, and the sample sizes used for each π and Tajima's D calculation in the relevant figure legends.

Response 2:

We have now revised both the Materials and Methods section and the relevant figure legends to provide the additional technical details requested. To further ensure reproducibility, all scripts used for read mapping, variant calling, and variant filtering have also been deposited in the dedicated **"Variant-Calling"** subfolder of our GitHub repository.

We have first clarified the SNP-calling strategy in the Materials and Methods (lines 1015–1022, under the subheading 'Salicornia population genomics'). For the genome-wide population genomic analyses, including nucleotide diversity (π), Tajima's D, sliding-window F_{ST} , and the chromosome-wide haplotype analyses presented in **Supplementary Figs. 17–19**, sequence reads from both *Salicornia* sp. France and *S. procumbens* were mapped to the chromosome-scale *Salicornia* sp. France reference genome, and variants were called independently for the A and C subgenomes. Prior to variant calling, raw reads were quality-trimmed (Phred score ≥ 15), aligned to the reference genome, and only uniquely mapped concordant read pairs were retained. Per-sample sequencing coverage was evaluated during quality control and averaged

approximately 10–12x across the dataset. Because all retained samples exhibited adequate coverage, no additional per-sample depth filter was applied. Variants were subsequently filtered using bcftools (QUAL ≥ 30 , INFO/DP ≥ 20 , allele frequency $0.01 < AF < 0.99$), followed by filtering for minor allele frequency (MAF ≥ 0.01), missing data $\leq 20\%$, and heterozygosity $\leq 20\%$.

Population-genetic statistics (π and Tajima's D) were calculated from the same filtered SNP dataset after separating samples by population (*Salicornia* sp. France, $n = 71$; *S. procumbens*, $n = 12$; defined now in the relevant figure legends of **Supplementary Fig. 17-19**). No additional variant filtering was performed after splitting the VCF; therefore, all analyses were based on the same set of high-confidence loci, differing only in the samples assigned to each population.

To evaluate the potential effect of unequal sample sizes, we repeated the π and Tajima's D analyses after randomly down-sampling the *Salicornia* sp. France population to 12 individuals, matching the sample size of *S. procumbens*. The sampled individuals, input files, and scripts used for the down-sampling analysis are provided in the GitHub repository. The contrasting population-genetic patterns were consistently reproduced, indicating that the observed signals were not driven by sample-size imbalance, as described in the Results section. Similarly, the sliding-window F_{ST} analysis was performed using balanced population sizes (12 individuals per population/species; defined now in the figure legend of **Supplementary Fig. 20**).

For Extended Data Fig. 10, we performed an independent validation by remapping reads from both populations to the diploid *S. ramosissima* reference genome and recalling variants within the OSCA candidate region. We have updated the figure legend to specify the reference genome used, the sample sizes ($n = 71$ and $n = 12$), the variant filtering criteria, and the quality-control summaries for the OSCA region, including the mean per-sample missingness and observed heterozygosity for each population. This independent analysis confirms that the observed haplotype differentiation is robust and is not dependent on the *Salicornia* sp. France reference assembly.

3. I would also encourage the authors to explain why HISAT2 was used for WGS read mapping, as this is not the most conventional choice for genomic resequencing data. For example, are there any studies or references to support that HISAT2 is the better choice in this case?

Response 3:

We thank the reviewer for this valuable comment. Although HISAT2 is widely recognized as an RNA-seq aligner, it can also be used for genomic DNA alignment by disabling splice-aware alignment. Accordingly, all whole-genome sequencing (WGS) reads in this study were aligned using HISAT2 with the `--no-spliced-alignment` option, thereby restricting alignment to genomic DNA. We also specified the maximum paired-end fragment length (`-X 1000`) to match the insert-size distribution of our sequencing libraries.

We selected HISAT2 because its Hierarchical Graph FM (HGFM) indexing strategy provides fast and memory-efficient alignment while maintaining high mapping accuracy in genomes with substantial sequence diversity, repetitive content, and structural variation, characteristics that

are common in *Salicornia*. Importantly, we empirically evaluated the accuracy of this mapping strategy using an in silico combined octaploid reference comprising the A and C subgenomes of *Salicornia* sp. France together with the B and D subgenomes of *S. bigelovii*. As shown in Supplementary Fig. 15, reads from representative accessions mapped almost exclusively to their expected subgenomes, whereas cross-subgenome and homeologous mapping to non-target subgenomes remained negligible ($\leq 2\%$). These results demonstrate high mapping fidelity and confirm that the alignment strategy provides robust subgenome discrimination for downstream analyses.

Our choice of HISAT2 is also supported by its successful application in previous genomic studies, including chromosome-scale genome assembly, whole-genome resequencing, and large-scale wheat skim-sequencing (Shrestha et al., 2023; Ahmed et al., 2023), where it was similarly used for genomic DNA alignment.

Overall, both the empirical validation presented in Supplementary Fig. 15 and previous applications demonstrate that the selected HISAT2 configuration provides accurate and efficient alignment for the genomic analyses performed in this study.

Shrestha, S., Koo, D.-H., Evers, B., Wu, S., Walkowiak, S., Hucl, P., Pozniak, C., Fritz, A., & Poland, J. (2023). Wheat doubled haploids have a marked prevalence of chromosomal aberrations. *The Plant Genome*, 16, e20309. <https://doi.org/10.1002/tpg2.20309>.

H. I. Ahmed, M. Heuberger, A. Schoen, D.-H. Koo, J. Quiroz-Chavez, L. Adhikari, et al. *Nature* 2023 Vol. 620 Issue 7975 Pages 830-838. DOI: 10.1038/s41586-023-06389-7.

Reviewer #1 (Remarks on code availability):

The codes used for the revised and newly added analyses appear to be missing. Given that several key conclusions now rely on additional analyses, including down-sampling, nucleotide diversity scans, remapping to an independent reference, and local haplotype visualization, the corresponding scripts and parameter settings should be made available.

Response:

We thank the reviewer for this valuable suggestion. We have substantially expanded and reorganized our GitHub repository (halophyte-salicornia-genomics) to ensure that all revised and newly added analyses are fully reproducible. The repository now contains dedicated subfolders (Variant-Calling, TajimasD-nucleotideDiversity_Fst, Genome-Coverage-Graphs, Haplotype-Analysis, and Phylogeny-PCA), each with a detailed README describing the workflow, required input files, software requirements, key parameters, execution order, and expected outputs.

We have added the scripts and processed input files required to reproduce the down-sampling experiment, nucleotide diversity (π), Tajima's D, sliding-window F_{ST} , remapping to the independent *S. ramosissima* reference genome, local haplotype visualization, genome coverage analysis, phylogenetic reconstruction and PCA analyses. Because the complete VCF files are too large for GitHub, they have been deposited in the Dryad repository and are linked through the GitHub documentation. Links to the GitHub repository are also provided in the manuscript (Lines 1288–1293).

In addition, the code currently provided on GitHub would benefit from clearer documentation and annotation. At present, it is difficult for other researchers to understand the required inputs, software versions, key parameters, expected outputs, and the order in which the scripts should be run. I recommend that the authors provide a more complete and better annotated workflow, including example commands or a README file, so that the revised analyses can be reproduced by readers.

Response 4:

We thank the reviewer for this helpful recommendation. We have comprehensively reorganized and documented the GitHub repository (halophyte-salicornia-genomics) to improve its clarity, accessibility, and reproducibility.

Specifically, we have:

- i. Reorganized the repository into dedicated analysis-specific subfolders (Variant-Calling, TajimasD-nucleotideDiversity_Fst, Genome-Coverage-Graphs, Haplotype-Analysis, and Phylogeny-PCA).
- ii. Added comprehensive README files for the main repository and for each analysis subfolder, describing the purpose of the analysis, required input files, software dependencies, key parameters, execution order, and expected outputs.
- iii. Updated and annotated all analysis scripts with consistent naming and workflow organization.
- iv. Included all required processed input files used for the revised analyses whenever practical.
- v. Organized the scripts and output files to correspond directly with the relevant figures and supplementary figures presented in the manuscript, making it straightforward for readers to identify and reproduce each analysis.

- vi. These revisions substantially improve the transparency and reproducibility of the study and provide readers with a complete, well-documented workflow for reproducing all analyses presented in the revised manuscript.

Reviewer #3 (Remarks to the Author):

The authors have revised the manuscript and adequately addressed my concerns. I have no further comments.

Response 5: The authors thank Reviewer #3 for the time taken to review this manuscript.