

Centromeric retention and recurrent introgression preserve adaptive ancestry across ploidy barriers in common reed

Lele Liu

liulele@sdu.edu.cn

Shandong University

Cui Wang

<https://orcid.org/0000-0003-2916-3922>

Wenyi Sheng

Shandong University

Wenbo Luo

University of Helsinki

Lele Lin

Ecology and Nature Conservation Institute, Chinese Academy of Forestry

Maja Ilievska

University of Helsinki

Huijia Song

Natural History Museum of China

Yuhui Wang

Shandong University

Weihua Guo

Shandong University

Article

Keywords: Polyploid evolution, Cross-ploidy ancestry, B chromosome, Salinity adaptation, Centromere evolution

Posted Date: September 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10599950/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

**Centromeric retention and recurrent introgression preserve adaptive ancestry across
ploidy barriers in common reed**

Cui Wang ^{1,2}, Wenyi Sheng ¹, Wenbo Luo², Lele Lin ³, Maja Ilievska², Huijia Song ⁴, Yuhui Wang ⁵, Weihua Guo ^{1,*}, Lele Liu ^{1,*}

¹ School of Life Sciences, Shandong University, 72 Binhai Road, Qingdao 266237, China

² Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

³ Ecology and Nature Conservation Institute, Chinese Academy of Forestry, No.1 Dongxiaofu, Xiangshan Road, Haidian district, Beijing 100091, China

⁴ Natural History Museum of China, 126 Tianqiao South Street, Beijing 100050, China

⁵ National Glycoengineering Research Center, Shandong University, Qingdao 266237, China

Corresponding author: Lele Liu (liulele@sdu.edu.cn); Weihua Guo (whguo_sdu@163.com)

Data availability Statement

The RNAseq data for the 18 samples were deposited in NCBI SRA database, with accession number PRJNA1066047.

Funding statement

This work is supported by the National Natural Science Foundation of China (No. 32100304; 32470388; U22A20558; 32271588) and the Natural Science Foundation of Shandong Province (ZR2021QC119).

Abstract

Recombination suppression is widely regarded as a barrier to genetic exchange, yet whether it can instead preserve shared ancestry across ploidy barriers and subsequently facilitate ecological adaptation has remained unknown. Combining chromosome-scale population genomics and transcriptomics from 126 individuals spanning multiple ploidy levels and salinity ecotypes of the common reed (*Phragmites australis*), we show that ancestry shared between an octoploid and a tetraploid lineage is disproportionately concentrated in centromeric and pericentromeric regions across the genome, where large shared haploblocks persist despite extensive recombination-driven erosion elsewhere and are enriched for functionally coherent gene clusters. At one locus, this retained haploblock has been independently recruited by selective sweeps in geographically isolated coastal and inland saline populations, linking centromeric ancestry retention to repeated ecological adaptation. Remarkably, the same recombination-suppressed genomic architecture produces the opposite evolutionary outcome at other loci, where haplotypic inversions promote rapid, lineage-specific centromere divergence rather than ancestry retention. Together, these findings reframe the classical centromere paradox, in which centromeric sequences turn over rapidly despite conserved centromere function, by showing that recombination suppression can drive either long-term ancestry retention or accelerated centromeric divergence, depending on local structural context. Our findings identify a centromeric region that has been repeatedly recruited for ecological adaptation across ploidy barriers, and suggest that centromeres more broadly may act as recurrent reservoirs of adaptive variation wherever polyploidization and hybridization coincide.

Key words: Polyploid evolution, Cross-ploidy ancestry, B chromosome, Salinity adaptation, Centromere evolution

1 Introduction

2 Polyploidization is a pervasive feature of plant evolution, with recurrent whole-genome
3 duplication events shaping the diversification of nearly all major angiosperm lineages¹⁻⁴.
4 Compared with their diploid progenitors, polyploids frequently display enhanced tolerance to
5 environmental stresses such as drought, salinity, temperature extremes, and nutrient deficiency
6 ⁵, and often colonize a broader range of ecological niches. Ploidy differences have long been
7 considered strong barriers to gene flow, owing to meiotic irregularities and reduced fertility in
8 interploidy hybrids. However, increasing evidence suggests that gene flow across ploidy
9 barriers is more common than previously appreciated, particularly in recently diverged
10 polyploid complexes⁶. As populations adapt to heterogeneous environments, locally
11 advantageous alleles may spread through adaptive introgression, whereby hybridization
12 enables the transfer of adaptive genetic variation between populations or species and this
13 process has been increasingly recognized as an important mechanism underlying plant
14 adaptation and evolution⁷. However, the evolutionary mechanisms underlying the enhanced
15 fitness of polyploids remain poorly understood.

16 Centromeres are essential for kinetochore assembly and faithful chromosome segregation,
17 yet despite this functional conservation across eukaryotes, centromeric sequences and repeat
18 elements evolve remarkably rapidly, a phenomenon termed the centromere paradox^{8,9}.
19 Centromeric divergence between hybridizing lineages can disrupt chromosome segregation
20 and drive reproductive isolation¹⁰, making centromeres a proposed hotspot of genomic
21 incompatibility. At the same time, recombination suppression in centromeric and
22 pericentromeric regions may preserve large ancestral haploblocks over long evolutionary
23 timescales, despite also facilitating the accumulation of linked deleterious mutations¹¹. Because
24 recombination is suppressed in pericentromeric regions, any ancestral centromeric segment
25 may persist intact across generations, potentially carrying adaptive variation across lineage
26 boundaries. Cross-ploidy ancestry enriched in pericentromeric regions has been documented
27 in *Alnus glutinosa*¹², and pericentromeric introgression accompanied by epigenetic remodeling
28 has been reported in cotton and wheat, suggesting this may be a broader feature of hybridizing
29 polyploid complexes. Nevertheless, the extent to which cross-ploidy ancestry facilitates
30 ecological adaptation, which genomic regions are permissive to interploidy gene flow, and how
31 retained variants are subsequently co-opted under natural selection have received little
32 empirical attention.

The common reed, *Phragmites australis*, offers an exceptional system in which to address these questions. As one of the most widely distributed grasses on Earth, *P. australis* encompasses lineages spanning tetraploid to octoploid ploidy levels¹³, and occupies habitats ranging from freshwater marshes and coastal saline wetlands to inland salt lakes and terrestrial grasslands^{14,15}. Multiple phylogenetically distinct lineages co-occur across Eurasia and have experienced repeated opportunities for hybridization and introgression, including across ploidy barriers¹⁶⁻²¹. The species has become a global model for studying biological invasion, local adaptation, and the genomic consequences of polyploidization^{22,23}. Crucially, the existence of well-characterized ecotypes adapted to contrasting salinity regimes²⁴, combined with a recently assembled chromosome-scale reference genome²⁵, makes *P. australis* uniquely suited for dissecting the interplay between cross-ploidy introgression and ecological adaptation at genomic resolution.

In China, two major haplotypes were discovered: haplotype O (denoted as CN tetraploid) primarily distributed in northern China, and haplotype P (denoted as AU octoploid) found in Korea and middle to southern China as well as Australia^{17-19,26,27}. Ploidy level, along with sexual (seed production) and asexual (rhizome growth) reproductive modes, are key traits contributing to the strong environmental tolerance of *Phragmites* species²⁸⁻³⁰. Population with higher genetic diversity, which may correlate with higher ploidy levels, have been found to be more tolerant to high salinity³¹. The distribution of ploidy levels appears geographically specific, and individuals with the same ploidy level may have formed from different polyploidization events³².

Here, we integrate whole-genome resequencing of 126 individuals spanning the major *Phragmites* lineages with transcriptomic profiling of 223 samples across ecotypes and salinity treatments, anchored to a chromosome-scale reference genome spanning all 25 chromosomes. Using this framework, we aim to characterize the genomic architecture across the *P. australis* polyploid complex, determine which chromosomal compartments tend to retain to interploidy gene flow, and establish whether retained ancestral segments have been co-opted for ecological adaptation in populations occupying contrasting salinity environments. By integrating selective sweep analyses with co-expression network reconstruction, we seek to establish a mechanistic link between cross-ploidy ancestry, centromere evolution, and parallel adaptation, addressing whether centromere-associated ancestry represents a general route by which polyploids access adaptive genetic variation across ploidy barriers.

Results

Variant calling and mapping accuracies

Using whole-genome sequencing data, a total of 126 samples, including *P. japonicus*, *P. karka*, *P. mauritanus*, and several geographic lineages of *P. australis* were mapped to the reference genome, yielding 277,829,484 variants across the 25 chromosomes (**Fig. 1a**; **Supplementary Table S1, S2**). After extracting SNPs, applying quality filtering, and pruning for linkage disequilibrium (10-kb window, 10-SNP step), we retained 232,589 SNP sites. These sites were used to construct a nuclear phylogenetic tree and to perform principal component analysis (PCA) for the *P. australis* lineages. A total of 1,584.53 Gb of clean RNA-seq data was obtained from 223 samples, averaging 5.88 Gb per sample. Reads from all 223 libraries were mapped to the reference genome with high accuracy, with unique mapping rates ranging from 85.94% to 93.96% (**Supplementary Table S1**).

Genomic, chromosomal and ecological phylogeny of *P. australis* lineages

As revealed by Lambertini *et al.* (2006)³³, *P. mauritanus* and *P. karka* are the most distantly related species to *P. australis* and were therefore selected as outgroup taxa to root the phylogeny. The resulting tree placed the hexaploid Gulf of Mexico Land-type lineage (USland) at the base of *P. australis*, while the population traditionally regarded as native North American *P. australis* clustered instead with *P. japonicus*, separate from all other *P. australis* lineages, and diverged early from the rest of the phylogeny¹⁸ (**Fig. 1b**; see **Supplementary Note 1** for discussion). This pattern indicates that the "native North American lineage" is more closely related to *P. japonicus* than to core *P. australis*, suggesting that these populations may in fact represent *P. japonicus* or a *P. japonicus*-derived lineage rather than true native *P. australis*, a possibility with direct implications for the taxonomic status of North American *Phragmites* populations historically classified as native *P. australis*.

Within *P. australis* proper, the octoploid AU lineage from southern China diverged earliest among the remaining lineages, followed by the divergence of the Med and EU tetraploid lineages from the northern Chinese populations (**Fig. 1b**). Ploidy level corresponded closely with phylogenetic position: the USland lineage is hexaploid, the AU lineage is octoploid, and the CN, EU, and Med lineages are all tetraploid, with these ploidy differences broadly tracking the major phylogenetic splits. Although the genome-wide phylogeny placed USland at the base of *P. australis*, per-chromosome trees revealed its clustering with the Med lineage on chromosomes 2, 19, and 25 (**Supplementary Fig. S1**), suggesting localized reticulate evolution or incomplete lineage sorting at these regions.

PCA results were broadly consistent with the phylogeny. *P. mauritanus*, *P. karka*, and the putatively hybrid USland lineage separated from the core *P. australis* groups along PC1. The

North American lineage clustered with *P. japonicus* and was likewise separated from *P. australis* along PC1, while positioned closer to the AU octoploid and CN tetraploid lineages along PC2 (**Fig. 1c; Supplementary Fig. S2**), consistent with the phylogenetic placement above. The four core *P. australis* lineages, AU octoploid, CN tetraploid, EU, and Med, were well separated along PC2, which accounted for 6.93% of the total variance.

Genetic admixture analysis of *P. australis* tetraploids identified four genetically distinct clusters corresponding to the seawater ecotype, the salt-lake ecotype, and the European (EU) and Mediterranean (Med) lineages. The freshwater ecotype showed no independent ancestry of its own, instead comprising entirely admixed genotypes derived from the seawater and salt-lake clusters, consistent with a hybrid origin or ongoing gene flow between these two ecotypes (**Supplementary Fig. S3**). Interestingly, four individuals sampled from the Tibetan Plateau exhibited mixed ancestry from the CN salt-lake, EU, and Med clusters, indicating that this region constitutes a secondary contact zone among otherwise geographically and genetically distinct lineages (**Fig. 1d**). Together, the widespread distribution of ploidy levels and the presence of such secondary contact zones provided recurrent opportunities for cross-ploidy introgression across the *P. australis* complex.

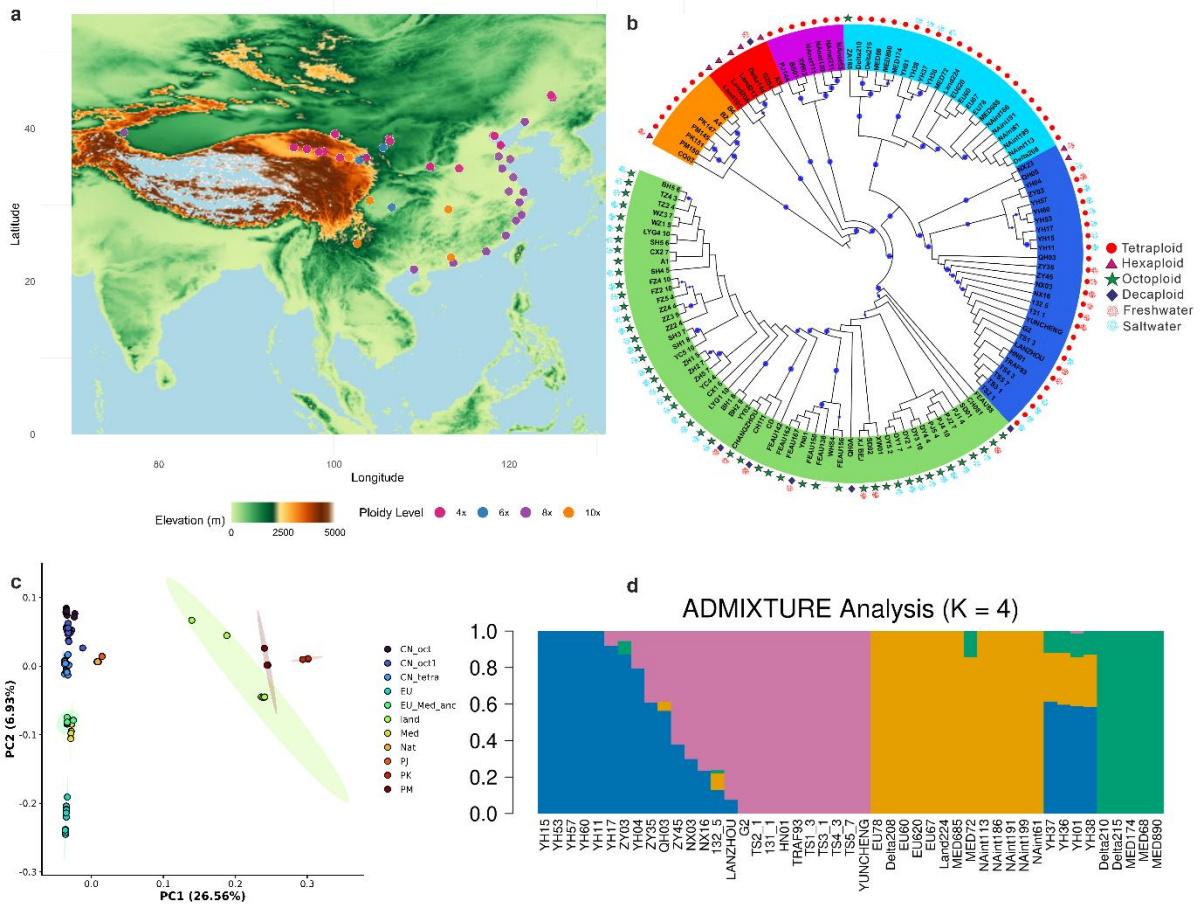


Figure 1. Phylogeographic diversity and population structure of *Phragmites australis* and

related species. (a) Geographic distribution of newly sampled *Phragmites australis* across East Asia. Each point represents a sampling locality, colored by ploidy level. **(b)** Maximum-likelihood phylogenetic tree inferred from genome-wide SNPs across all 25 chromosomes. Tip label colors denote genetic lineages: orange, *P. mauritanus*/*P. karka*; red, the Gulf of Mexico Land-type lineage; purple, *P. japonicus*; light blue, the Mediterranean/European tetraploid lineage of *P. australis*; dark blue, the Chinese tetraploid lineage of *P. australis*; green, the Australian octoploid lineage of *P. australis*. Outer ring symbols indicate ploidy level; red and blue spiral symbols denote the freshwater and saltwater ecotypes of *P. australis*, respectively. **(c)** Principal component analysis (PCA) of all *Phragmites* species and lineages. PC1 (26.56% of variance) separates major species, while PC2 (6.93% of variance) resolves intraspecific lineages. The increasing genetic distance of Australian octoploids from Mediterranean/European lineages along PC2, with Australian octoploids clustering closer to Chinese tetraploids, is consistent with Chinese tetraploids representing a potential ancestral lineage. **(d)** ADMIXTURE analysis of *P. australis* tetraploids ($K = 4$), illustrating genetic differentiation and potential gene flow among lineages and ecotypes. Each vertical bar represents one individual, and bar colors represent ancestry proportions from each inferred genetic cluster.

Centromeric identification

Centromeric regions were initially predicted using CentIER, which integrates k-mer frequency, tandem repeat (TR), long terminal repeat (LTR), and Hi-C data (**Supplementary Fig. S4**). Because CentIER frequently over-estimated the extent of centromeric regions and occasionally misclassified centromere position, all predictions were manually refined using Hi-C contact heatmaps, in which centromeres were identified as regions exhibiting reduced contact frequencies between genomic bins located on opposite chromosome arms relative to contacts among bins within the same chromosome arms³⁴, together with the distribution of gene density and tandem repeat density (**Supplementary Fig. S5**). Following this correction, centromeric regions were confidently identified in 24 of the 25 chromosomes; PaChr6D could not be predicted by CentIER directly and was instead located manually from the Hi-C contact map and flanking gene/repeat density, while the B chromosome PaChr24B could not be resolved by either approach, likely owing to its high heterochromatin content. Manual correction also revised the centromere position initially assigned to PaChr18A, which CentIER had misclassified as acrocentric; after correction, PaChr18A was reassigned to the metacentric/submetacentric group along with most other chromosomes. Based on the corrected

centromere positions, all chromosomes are metacentric or submetacentric with the sole exception of telocentric PaChr20D (**Supplementary Table S3; Supplementary Fig. S4, S5**). PaChr3D and PaChr10A each yielded two candidate regions, suggesting possible metacentric configurations that were not fully resolved. GO enrichment analysis of these pericentromeric regions uncovered several functionally coherent gene clusters, including two tandem "vesicle targeting" genes on PaChr1A, an array of five tandem "cell fate determination" genes on PaChr10A, and an array of nine tandem genes associated with "symbiont-induced defense-related programmed cell death" on PaChr21D. A haplotypic inversion spanning the centromeric region of PaChr22A has previously been identified in the reference genome assembly²⁵, coinciding with the centromere divergence described below.

Genome-wide introgression reveals localized retention in centromeres

To distinguish incomplete lineage sorting from introgression, we first quantified genome-wide gene flow using D-statistics under the topology (((CN, EU_Med), AU), *P. japonicus*). Genome-wide analyses yielded a significantly negative D statistic ($D = -0.0651$, $Z = -24.64$, $P < 0.001$), indicating an excess of BABA over ABBA patterns and demonstrating that AU shares significantly more derived alleles with CN than with EU_Med, consistent with historical introgression between AU and CN.

This pattern was largely consistent across chromosomes. After multiple-testing correction, 17 of the 25 chromosomes showed significantly negative D values (adjusted $P < 0.05$; $|Z| > 3$), with the strongest signal observed on PaChr12D ($D = -0.120$, $Z = -9.77$, adjusted $P = 1.25 \times 10^{-21}$) (**Supplementary Table S4**). Seven chromosomes (PaChr4A, PaChr6D, PaChr7A, PaChr10A, PaChr13D, PaChr17A, and PaChr20D) showed no significant chromosome-wide deviation from zero. Sliding-window analyses, defining significant introgression as regions in which $|fdM|$ exceeded 0.2 across more than three consecutive windows, further revealed localized introgression between AU and EU_Med within centromeric or pericentromeric regions of PaChr4A, PaChr6D, PaChr7A, PaChr13D, and PaChr17A (**Supplementary Table S5, S6**). In addition, a localized AU–EU_Med introgression signal was detected at 2.2–2.6 Mb on PaChr10A, despite the absence of a significant chromosome-wide D statistic. These results indicate that chromosome-wide analyses underestimate fine-scale heterogeneity in introgression and suggest that introgressed haplotypes are preferentially retained in centromeric regions.

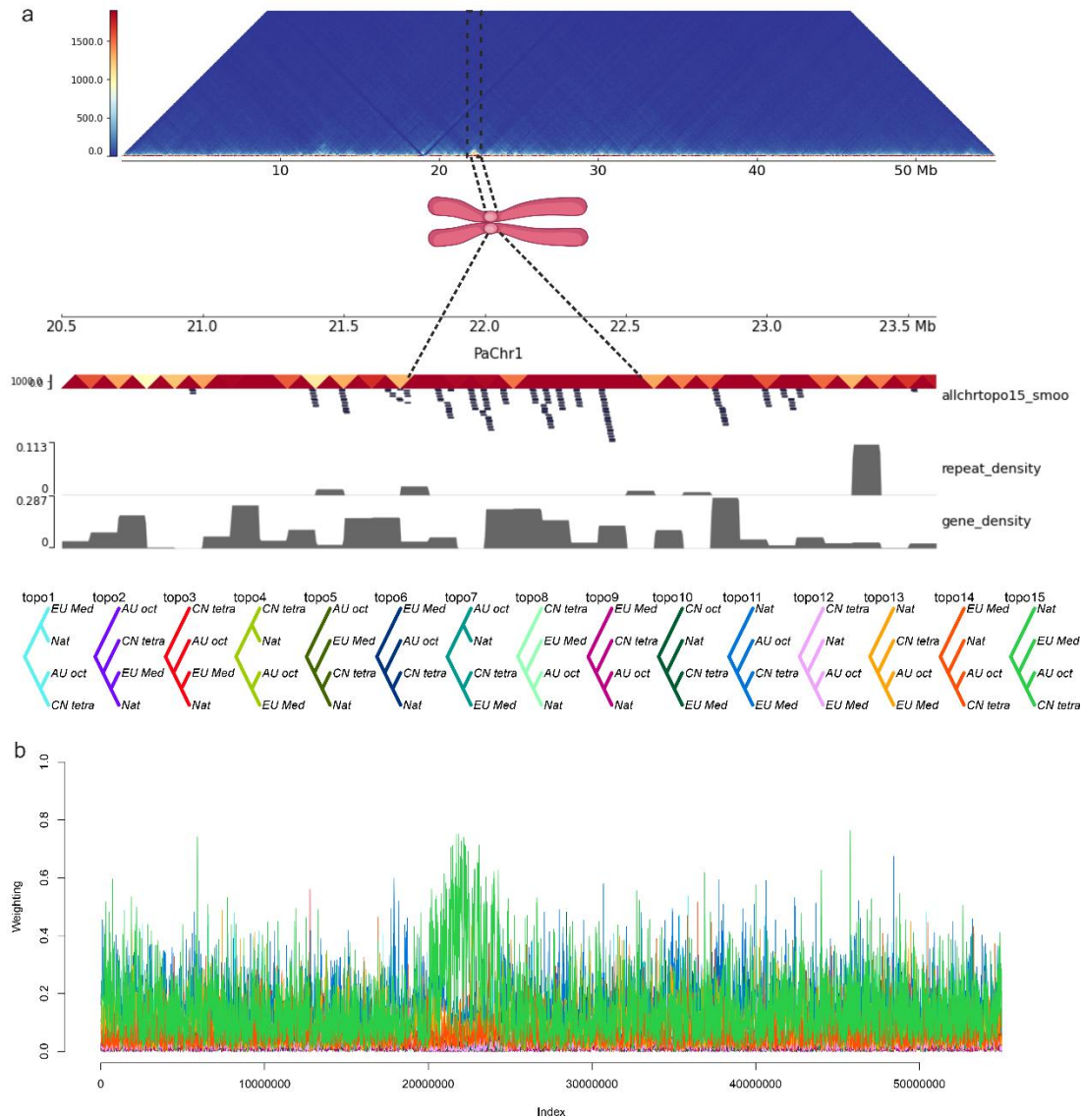


Figure 2. (a) **Hi-C contact map of chromosome PaChr1A in *P. australis*.** Intra-chromosomal Hi-C interaction map of PaChr1A, shown as a triangular heatmap generated with pyGenomeTracks at 100 kb bin resolution. The heatmap is displayed from the main diagonal up to a depth of one-quarter of the chromosome length. The color bar indicates the Hi-C contact counts between pairs of 100 kb genomic bins, with warmer/darker colors denoting more frequent contacts; values are shown on a linear scale. The genomic coordinate axis (in Mb) is shown at the bottom. The region between the two dashed lines indicates the centromeric region. Below, a zoomed-in Hi-C contact heatmap of this centromeric region is shown together with the density of tandem repeats identified by the TRASH pipeline and gene density along the same interval. (b) Topology weighting along chromosome PaChr1A in *P. australis*. Topology weightings were estimated using Twisst2 from genotype data, with up to 512 subtrees sampled per genealogy. Four groups were defined for the analysis: Nat, EU_Med, CN_tetra, and AU_oct. Weightings were smoothed using a 20 kb span with 1 kb spacing and are plotted along

the chromosome (in Mb). At each position, the stacked weightings sum to 1 and represent the relative frequency with which each of the 15 sampled topologies is supported by the local genealogies; the corresponding topologies are shown above. Higher weighting for a given topology indicates that genealogies matching that topology predominate in that genomic region.

Topology weighting localizes introgressed haploblocks to centromeric regions

To determine whether the introgressed haplotypes identified by D-statistics were randomly distributed across chromosomes or preferentially retained within specific genomic compartments, we next examined genome-wide topology weighting using Twisst2. Two lineage configurations were evaluated: Nat, EU_Med, CN, and AU to resolve relationships among the tetraploid and octoploid *P. australis* lineages, and *P. mauritanus*, EU_Med, *P. japonicus*, and the USland lineage to investigate the hybrid origin of USland.

Across the genome, the predominant topology grouped AU with CN (((CN, AU), EU_Med), Nat), in agreement with the genome-wide D-statistics supporting historical introgression between these two lineages. Elevated support for this topology was detected on PaChr1A, 3, 4, 6, 8, 10, 12–16, 18, and 20–25 (**Supplementary Fig. S6**), with the strongest weighting (weight higher than 0.15) concentrated in centromeric and pericentromeric regions (**Supplementary Table S7**). The same topology extended across the entirety of PaChr2D, PaChr5A, PaChr7A, PaChr9A and PaChr11D, as well as the 5.4–6.5 Mb interval of PaChr17A, indicating that centromeric regions preferentially retain the ancestral CN–AU relationship.

Despite this dominant pattern, several centromeres displayed localized phylogenetic discordance. Topologies supporting an AU–EU_Med sister relationship were enriched in the centromeric regions of PaChr3D, PaChr6D, PaChr9A, PaChr10A, PaChr14A, PaChr15A, and PaChr19D, as well as the 6.5–9.8 Mb interval of PaChr17A. The AU–EU_Med topologies on PaChr6D, PaChr10A and PaChr17A coincide with the localized AU–EU_Med introgression signals identified by sliding-window D-statistics, providing independent support for recent introgression in these regions. Additional AU–EU_Med topologies that were not accompanied by significant *fdM* values may instead reflect older introgression events or incomplete lineage sorting that are not detectable using allele-frequency-based statistics.

The mosaic arrangement of alternative topologies along PaChr17A is particularly striking. The transition from a CN–AU topology (5.4–6.5 Mb) to an AU–EU_Med topology (6.5–9.8 Mb) suggests chromosome breakage and refusion during the whole-genome duplication that gave rise to the AU lineage.

Most centromeres therefore retained shared ancestry among tetraploid lineages. Two chromosomes, however, departed from this general pattern. On PaChr22A, the AU centromere was highly divergent from those of the other three tetraploid lineages, consistent with the emergence of a lineage-specific centromere. Conversely, on PaChr4A, most of the chromosome supported a sister relationship between EU_Med and CN, whereas the centromeric region (26–30 Mb) was highly divergent and specific to the EU_Med lineage, indicating an independently evolved centromere. Consistent with the sliding-window D-statistics, localized topology shifts on PaChr4A further corresponded to putative AU–CN introgression (25.0–27.75 Mb) and AU–EU_Med introgression (~30 Mb), highlighting the complex evolutionary history of this chromosome.

Under the second lineage configuration, topology weighting revealed nearly equal genomic contributions from EU_Med and *P. mauritanus* across almost all chromosomes of the USland lineage (**Supplementary Fig. S7**), in agreement with the admixture analysis and confirming its hybrid origin.

Selective sweeps detected in isolated CN tetraploid populations

We estimated pairwise F_{ST} and the Population Branch Statistic (PBS) among three ecotypes of the CN allotetraploid lineage. Because ANGSD models genotype likelihoods under a diploid assumption, we restricted these analyses to the allotetraploid lineage, whose two subgenomes are resolved in the genome assembly and therefore behave as effectively diploid at the subgenome level. This lineage comprised five individuals from the salt lake, five from seawater, and ten from freshwater. To identify candidate regions under selection, we applied an empirical threshold corresponding to the top 1% of genome-wide F_{ST} and PBS values.

Genetic differentiation between the salt-lake and sea populations was generally higher than in the other two pairwise comparisons, as reflected by elevated baseline F_{ST} across the genome (**Supplementary Fig. S8**). The salt-lake population also showed elevated PBS values genome-wide. Nucleotide diversity (π) in the salt-lake population was only slightly higher than in the sea population (**Fig. 3**, bottom panel; **Supplementary Fig. S9**), a modest difference that argues against a severe bottleneck as the sole driver of the elevated PBS signal and instead points toward a genome-wide history of selection distinct from the other two ecotypes. Against this background, we identified a prominent region on PaChr7A (23.8–28.8 Mb) showing elevated F_{ST} in both the freshwater-vs-sea and salt-lake-vs-sea comparisons, together with high PBS peaks specific to the salt-lake and sea populations and negative PBS values in the freshwater population (**Fig. 3**). This pattern is consistent with a strong, ecotype-specific selective sweep and suggests that the salt-lake and sea populations have undergone independent, parallel

adaptation within this interval. Coincidentally, this region corresponds to the pericentromeric region of PaChr7A and overlaps with segments introgressed from the EU tetraploid lineage into the AU octoploid genome, indicating that these introgressed centromeric regions are themselves under positive selection. A second region showing a similar signature was identified on PaChr10A (15.2–20.0 Mb), corresponding to its pericentromeric region and likewise overlapping with CN-derived segments in the AU octoploid; however, in this case no elevated PBS signal was observed in the salt-lake population, suggesting that selection at this locus may be specific to the sea ecotype.

Within PaChr7A, we identified 70 candidate genes under selection (**Supplementary Table S8**), including genes coding for Metal Tolerance Protein C4, ABC Transporter B Family Member 19, Calmodulin-binding Domain-containing Protein, WRKY Transcription Factor 11, bHLH Domain-containing Protein, Receptor-like Serine/Threonine-protein Kinase, Wax Synthase Isoform 1, Extensin-like Protein, and Pathogenesis-related Protein 1, among others. Four of these candidates are tandem duplicates encoding a Glycosyl Transferase Family 28 C-terminal Domain-containing Protein. GO enrichment analysis revealed that these four tandem duplicate genes, PaChr7A.1590, PaChr7A.1592, PaChr7A.1594, and PaChr7A.1596, are significantly enriched in the GO term "dolichol-linked oligosaccharide biosynthetic process." The ABC transporter (ABC Transporter B Family Member 19) has been functionally implicated in auxin transport in other systems³⁵.

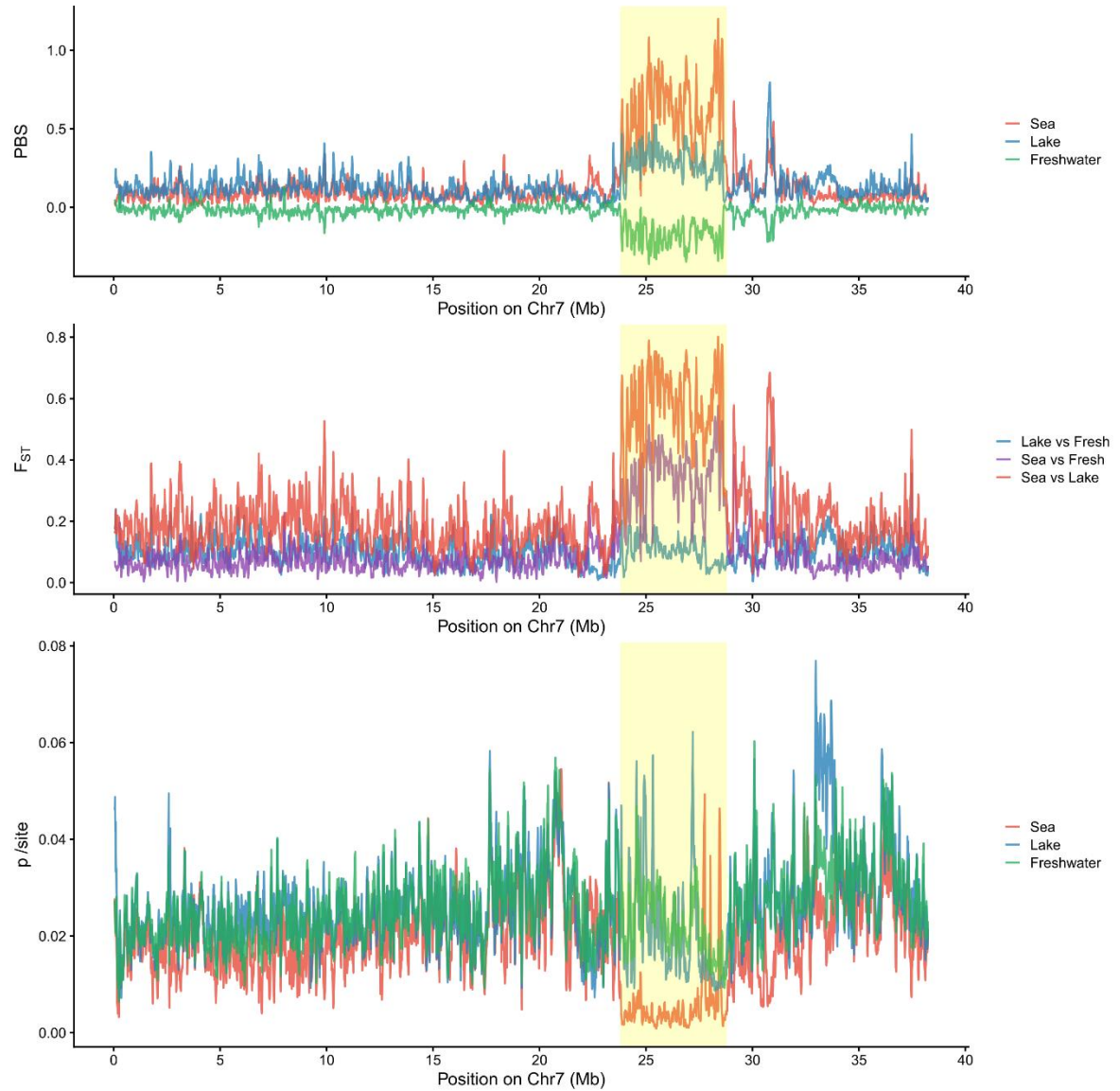


Figure 3. Genome-wide selection scan along chromosome PaChr7A in the CN allotetraploid lineage of *P. australis*. Population genetic statistics were estimated in 10 kb sliding windows along PaChr7A (0–40 Mb) for three ecotypes, salt-lake (Lake), seawater (Sea), and freshwater (Freshwater). (Top) Population Branch Statistic (PBS) for each ecotype, where elevated values indicate lineage-specific allele frequency differentiation consistent with population-specific selection. (Middle) Pairwise F_{ST} between ecotypes (Lake vs Fresh, Sea vs Fresh, and Sea vs Lake), measuring genetic differentiation between each pair of populations. (Bottom) Nucleotide diversity (π per site) for each ecotype, where reduced values may indicate loss of diversity due to selection. All statistics were computed from genotype likelihoods in ANGSD; the x-axis shows genomic position along PaChr7A in Mb.

The AU octoploid predominantly carries salt-lake-derived alleles

Two independent lines of evidence indicate that the adaptive haplotype under selection on PaChr7A originated from the CN salt-lake lineage, though the precise route by which it was acquired by the AU octoploid, whether through direct introgression from the salt-lake lineage, indirect transfer via the EU/Med lineage, and retention of shared ancestral polymorphism facilitated by low recombination in this pericentromeric region.

First, phylogenetic analysis based on whole-chromosome variants showed that the AU octoploid diverged from other genetic lineages immediately after the split between *P. australis* and *P. japonicus*, with all three CN ecotypes (sea, freshwater, and salt-lake) grouped within a single cluster (**Supplementary Fig. S1**). This contrasts sharply with the topology recovered specifically within the selected region on PaChr7A. Here, the sea population and a subset of freshwater individuals diverged earliest from the remaining *P. australis* populations, with only a small number of AU octoploid individuals grouped within this clade, likely reflecting limited, localized gene flow; the majority of AU octoploid individuals instead either formed a distinct cluster alongside, or grouped directly together with, the salt-lake allotetraploid population (**Fig. 4a**). The salt-lake cluster itself grouped together with the Mediterranean and European populations, raising the possibility that the EU/Med lineages acquired salinity tolerance through introgression from the salt lake lineage, and that the AU octoploid may in turn have acquired this haplotype either directly from the salt-lake lineage or indirectly through subsequent contact with EU/Med. Alternatively, given that this region lies within the pericentromeric interval where recombination is strongly suppressed, the shared topology among salt-lake, EU/Med, and AU could also reflect retention of a common ancestral haplotype that has simply not yet been broken down by recombination in any of these lineages, rather than a discrete introgression event. On PaChr10A, the sea population formed its own distinct cluster, accompanied by a small number of introgressed freshwater and octoploid individuals, while all remaining *P. australis* lineages grouped together (**Fig. 4b**); the source lineage is less unambiguous here than on PaChr7A, but the pattern is broadly consistent with the same general scenario. Within the sweep-region trees of both PaChr7A and PaChr10A, the majority of AU octoploid individuals grouped with the salt-lake cluster, with only a minority clustering with the sea population, indicating that the salt-lake lineage was the primary source of adaptive introgression at both loci and that sea-clustering octoploid individuals likely reflect secondary admixture. This is consistent with the ADMIXTURE-based finding that the freshwater population itself represents an admixture of salt-lake and sea ancestry rather than an independent ancestral state.

Second, we identified diagnostic SNPs that are completely fixed and exclusive to either the marine or salt-lake allotetraploid ecotypes, and examined their representation in AU octoploid individuals. Of 7,751 marine-specific SNPs, 61.6% (4,772) were detectable in the AU octoploid at a frequency ≥ 0.05 , but at a median frequency of only 0.09, indicating that marine-derived alleles, where present, occur at very low frequency. In striking contrast, 97.1% (23,665 of 24,362) of salt-lake-specific SNPs were detected in the AU octoploid, at a median frequency of 0.78, with 65.3% of these loci exceeding a frequency of 0.5 (**Fig. 4c,d**), although none reached complete fixation (frequency=1.0).

Taken together, these results establish that the adaptive haplotype under positive selection in the AU octoploid traces its origin to the CN salt-lake lineage. Whether this haplotype reached the AU octoploid through direct introgression, indirect transfer via the EU/Med lineage, or long-term retention under suppressed recombination, its persistence and subsequent positive selection highlight the salt-lake lineage as the ultimate source of this adaptive variation.

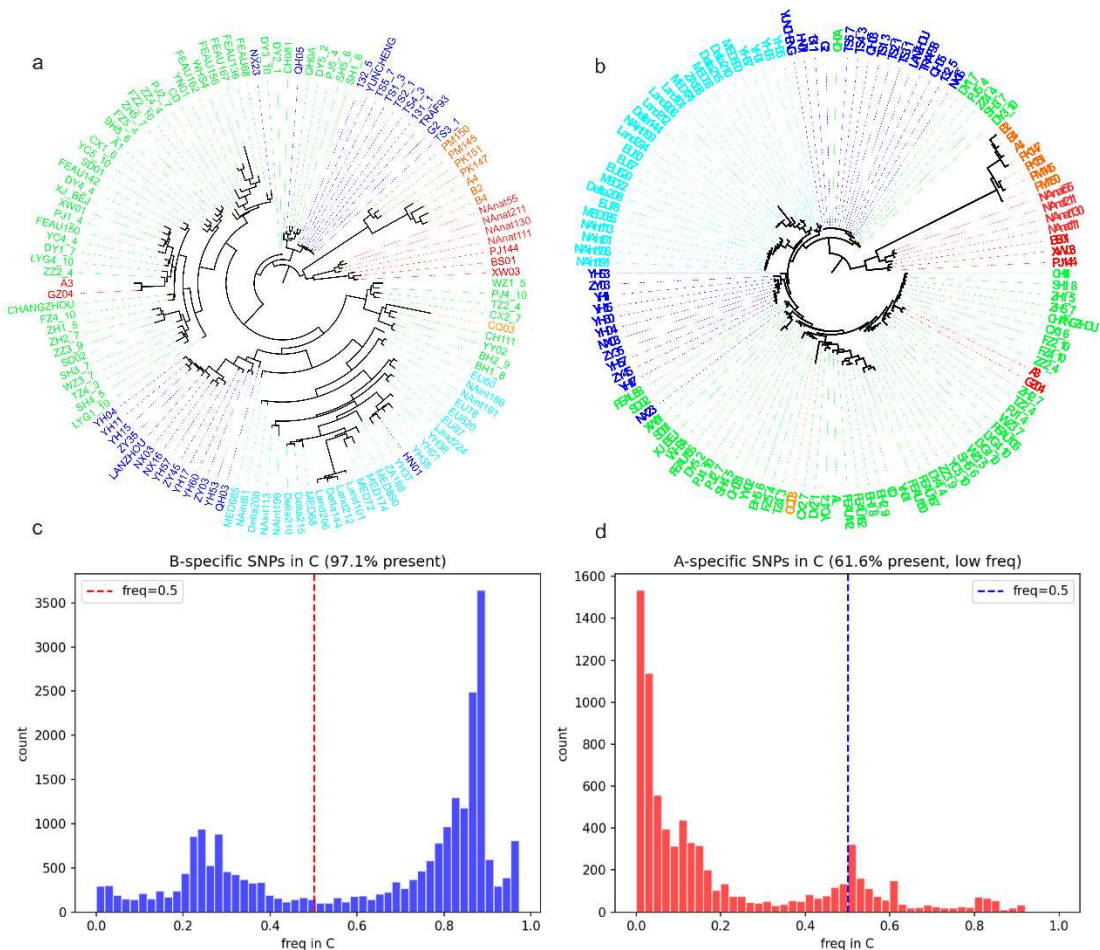


Figure 4. (a) Maximum-likelihood phylogeny of the selected region on PaChr7A (23.8–28.8 Mb). Tip colors denote lineage: red, *P. japonicus*; orange, *P. mauritanus*; dark blue, CN tetraploid; light blue, EU/Med lineage; green, AU octoploid. Samples whose names begin with

"TS" are seawater samples. **(b)** Maximum-likelihood phylogeny of the selected region on PaChr10A (15.2–20.0 Mb); colors are as in (a). **(c)** Frequencies of salt-lake-specific SNPs ($n = 24,362$) in the AU octoploid. **(d)** Frequencies of sea-specific SNPs ($n = 7,751$) in the AU octoploid. In (c) and (d), the dashed red line indicates a frequency of 0.5. The strong enrichment of salt-lake-specific alleles at high frequency suggests that the AU octoploid is derived primarily from the salt-lake lineage.

Gene expression in response to salt stress

Comparison between the control samples of sea and freshwater populations identified 32 B chromosome genes (gene name labeled on PaChr24B) as significantly upregulated among the 1831 upregulated genes in the sea population (**Supplementary Table S9; Fig. 5 b**), while the salt-lake population showed no such elevation relative to freshwater, with only 273 upregulated genes and 173 downregulated (**Supplementary Table S10**). A further 15 B chromosome genes were upregulated in the sea population compared to the salt-lake population (**Supplementary Table S11**), suggesting a sea population-specific role of B chromosome expression in salinity adaptation.

Gene expression profiles were compared across all samples under control and salt-treatment conditions. PCA revealed that population genetic background, rather than salt treatment, was the primary driver of expression variance (PC1: 19.7%; PC2: 9.4%), with AU octoploids separated from the tetraploid lineages along PC1 (**Fig. 5a**). To test whether the sea ecotype mounts a distinct transcriptional response to salt stress, we compared control and salt-treated samples within this population directly; strikingly, only 5 genes were significantly upregulated and 13 downregulated ($|\log_2\text{FC}| > 1$; **Supplementary Table S12**), indicating that the sea population shows little acute transcriptional response to salt treatment itself. This near-absence of a treatment-induced signature suggested that salinity tolerance in this ecotype may instead be encoded constitutively, prompting us to compare baseline expression between ecotypes under control conditions.

GO enrichment (biological process; Bonferroni $P < 0.05$) revealed a shared stress- and defense-response signature among genes up-regulated in the sea population against both reference populations, including defense response (to bacteria and fungi), salicylic acid and abscisic acid responses, and signal transduction (**Fig. 5c; Supplementary Table S13**). Relative to freshwater, up-regulated genes were additionally enriched for abiotic stress (hypoxia, oxidative, osmotic and salt stress), jasmonic acid and ethylene signalling, and nitrogen metabolism; relative to the salt-lake population, they were specifically enriched for

programmed cell death, the hypersensitive response, and inorganic ion transmembrane transport. Down-regulated genes were consistently associated with RNA metabolism and chloroplast/plastid RNA regulation across both comparisons, alongside reduced photosynthetic light-harvesting. No GO terms were enriched between the freshwater and salt-lake populations.

For the octoploid-versus-allotetraploid comparisons, the three ecotype contrasts shared only a small core of enriched GO terms (**Fig. 5d; Supplementary Table S13**). Among upregulated DEGs, the only term enriched across all three contrasts was the general "response to stimulus"; the strongest overlap was between the salt-lake and freshwater contrasts (43 shared terms), which together captured a defense- and stress-response signature (response to stress, defense response to bacteria and other organisms, response to biotic stimulus, signal transduction), whereas the seawater contrast shared almost none of its enrichment with the other two. Downregulated terms followed a similar pattern: biosynthetic process was enriched across all three contrasts, while translation was specifically suppressed in the seawater and salt-lake contrasts. Contrast-specific enrichments included cellular response to sulfur starvation and photosystem I light-harvesting (seawater, upregulated); suppression of DNA replication and genome maintenance (salt-lake, downregulated); and hypoxia, cold, and jasmonic acid responses (freshwater, upregulated).

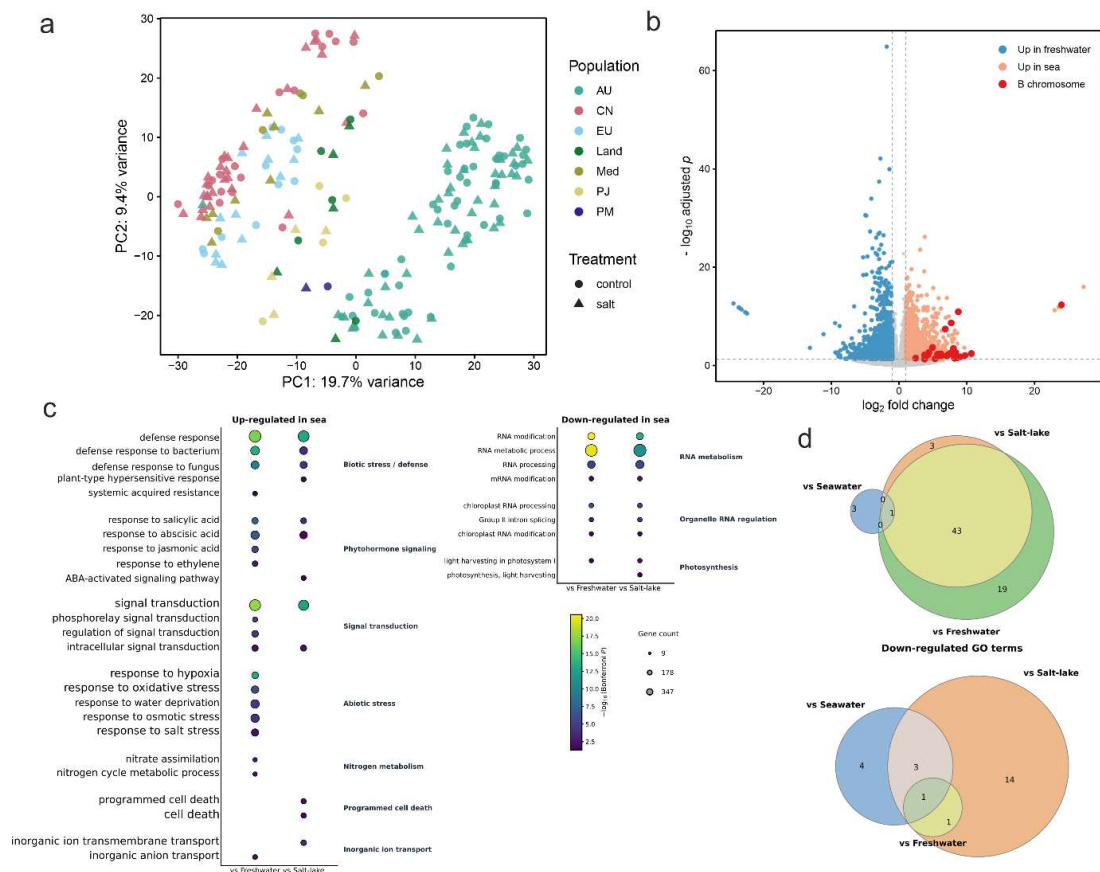


Figure 5. Transcriptomic divergence of *P. australis* across octoploids and saline ecotypes. **(a)** Principal component analysis (PCA) of regularized log-transformed (vsn) expression across all samples. Each point is one sample, coloured by population (AU, CN, EU, Land, Med, PJ, PM) and shaped by treatment (control vs salt). PC1 and PC2 explain 19.7% and 9.4% of the total variance, respectively. **(b)** Volcano plot of differentially expressed genes (DEGs) between the seawater and freshwater populations, showing $\log_2$ fold change against $-\log_{10}$ adjusted P. Significant DEGs ($|\log_2 \text{ fold change}| > 1$, adjusted $P < 0.05$) are shown in dark grey and non-significant genes in light grey; positive and negative $\log_2$ fold changes correspond to up-regulation in the sea and freshwater populations, respectively. Red points denote B-chromosome genes located on chromosome PaChr24B. **(c)** GO biological-process enrichment among genes differentially expressed in the sea population relative to the freshwater and salt-lake populations (Bonferroni $P < 0.05$). Up-regulated (left) and down-regulated (right) terms are shown separately; representative non-redundant terms are grouped by functional category (right-hand labels). Dot colour indicates $-\log_{10}(\text{Bonferroni } P)$ and dot size the number of differentially expressed genes annotated to each term. A dot is shown only where a term is significantly enriched in that comparison; absence indicates no significant enrichment. See Supplementary Table S9. **(d)** Venn diagrams of significantly enriched GO biological-process terms (Bonferroni $P < 0.05$) shared among the three octoploid-versus-allotetraploid ecotype comparisons (vs seawater, vs salt-lake, vs freshwater), shown separately for up-regulated (top) and down-regulated (bottom) terms. Numbers indicate the count of GO terms in each region. See Supplementary Table S10.

Co-expression network of stress-responsive genes among genetic lineages and salt treatments

WGCNA of 223 samples identified seven modules significantly correlated with species or lineage identity. Five modules with adequate sample size and statistical power were examined further: the greenyellow (positively correlated) and darkred (negatively correlated) modules in AU, the lightyellow module in the EU invasive population, the black module in the USland lineage, and the lightgreen module (**Supplementary Fig. S10**).

The lightgreen module was positively correlated with AU overall ($r = 0.37$) but negatively correlated with CN ($r = -0.24$). Closer inspection of sample identities showed that high eigengene values derived exclusively from Australian-origin individuals within the AU octoploid group, indicating that this module reflects a constitutive expression signature specific to the Australian lineage rather than a property of the AU group as a whole (**Supplementary Figure S11**). This module contained the three duplicated core genes under selection at

PaChr7A (PaChr7A.1592/1594/1596), co-expressed with 236 other genes, including tandem gene pairs (e.g., PaChr4A.853–854, PaChr9A.616–617, PaChr21D.470–471) and additional selected genes from the same sweep region (PaChr7A.1623/1627/1628). Rather than the canonical osmotic- or ion-transport-related processes typically associated with salt-stress response, this module was enriched for genome maintenance functions, including telomere maintenance, RNA-templated DNA biosynthesis, double-strand break repair via homologous recombination, the meiotic cell cycle, and nucleotide-excision repair (**Supplementary Table S14**). This indicates that the introgressed PaChr7A haplotype is embedded within a broader, lineage-specific co-expression network extending well beyond the selected locus itself, and that its integration into this network is centered on genome integrity rather than acute stress signaling. Details of the modules correlated with other genetic lineages and the gene expression network regulated by the B chromosome are provided in **Supplementary Note 2** and **Supplementary Tables S15** and **S16**.

Discussion

Cross-ploidy hybridization provides important opportunities for the exchange of genetic variation during polyploid evolution, yet why some genomic regions preserve this ancestry over evolutionary timescales while others are rapidly homogenized remains poorly understood^{36,37}. Here, we show that cross-ploidy ancestry in *P. australis* is strikingly concentrated within centromeric and pericentromeric regions, indicating that the genomic fate of exchanged ancestry is highly non-random. Because recombination across chromosome arms progressively erodes ancestry blocks through repeated reshuffling, much of the rest of the genome becomes homogenized over time, while centromeres retain large, coherent haploblocks regardless of whether that ancestry was inherited at polyploid formation or acquired later through introgression. Similar enrichment of introgressed or ancestrally retained variation in low-recombining regions has been reported in *Alnus glutinosa*, cotton, and wheat^{38,39}, suggesting this may be a general consequence of recombination suppression in polyploid genomes rather than a feature unique to *Phragmites*. We therefore propose a centromeric ancestry retention model (**Fig. 6**), in which centromeric regions act as long-term reservoirs of cross-ploidy ancestry. By suppressing recombination, these regions preserve large ancestral haploblocks that remain available for subsequent ecological adaptation long after their initial acquisition.

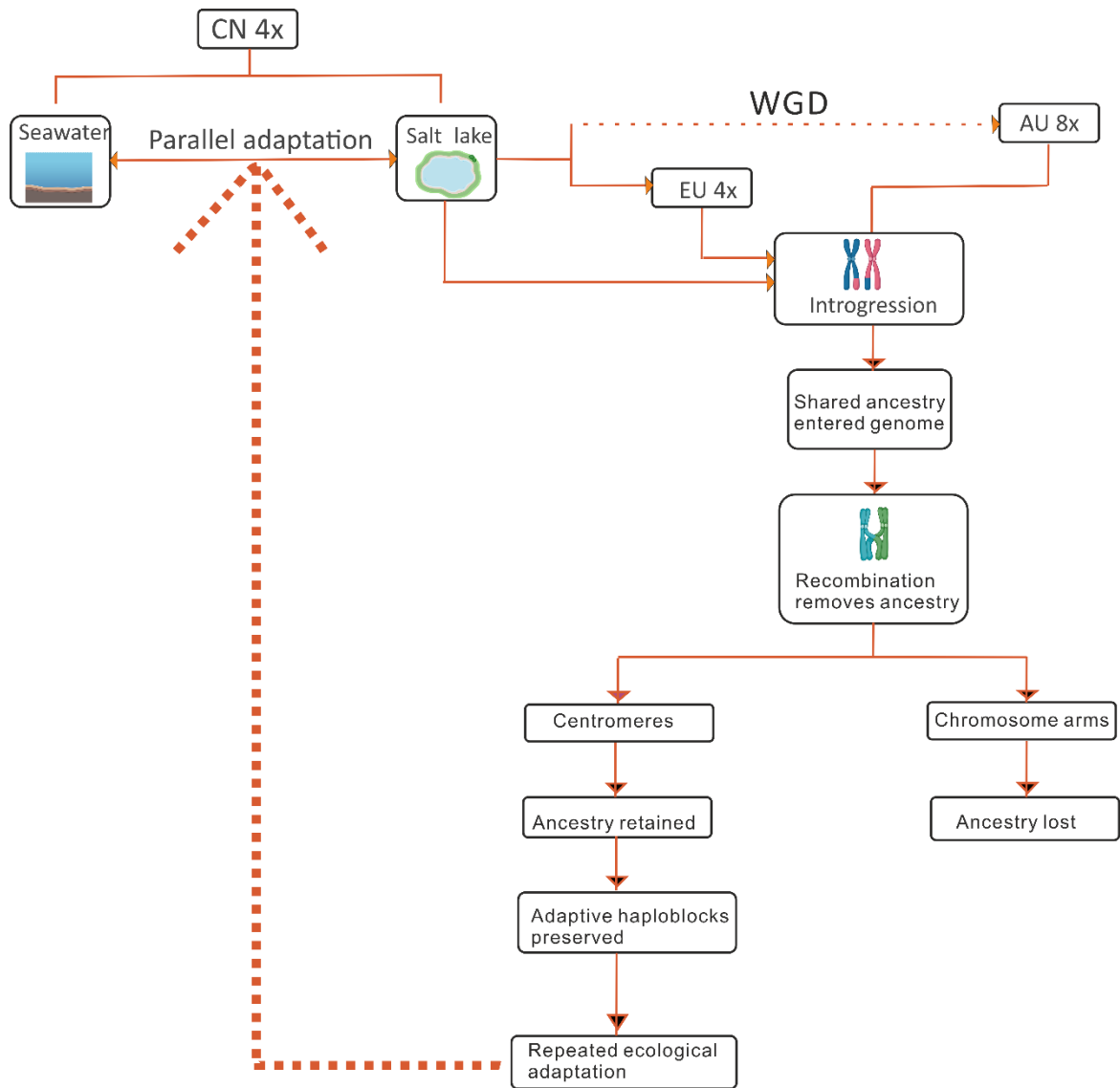


Figure 6. A centromeric ancestry retention model. Schematic model of how whole-genome duplication (WGD), introgression, and recombination suppression at centromeres enable repeated ecological adaptation in the common reed (*P. australis*) complex. An ancestral tetraploid lineage (CN 4x) diverged into seawater and salt lake ecotypes that convergently adapted to saline habitats (parallel adaptation). The salt lake lineage further gave rise to a European tetraploid (EU 4x) and, potentially via WGD, an Australian octoploid (AU 8x) lineage; secondary contact between them produced introgression, introducing shared ancestral haplotypes into the recipient genome. Subsequent recombination removed this introgressed ancestry from chromosome arms but not from centromeres, where recombination suppression preserved the introgressed haploblocks. Retention of these adaptive haploblocks at centromeres is proposed to underlie the repeated, convergent adaptation to saline environments observed across lineages. The dashed line showed the proposed route of origination of AU octoploids,

1 whereas the dot line showed the confirmed way of selected centromere functioned as adaptive
2 centromeres.

3
4 Although our data do not resolve the precise timing of ancestry acquisition for every
5 centromeric region, the consistent enrichment of both inherited and introgressed ancestry
6 within these regions indicates that recombination suppression governs their long-term
7 persistence regardless of origin. This interpretation is consistent with observations from diverse
8 systems, including the preferential retention of archaic hominin ancestry in modern human
9 centromeres⁴⁰, suggesting that low-recombination genomic compartments can preserve ancient
10 genetic variation over evolutionary timescales. Retention alone, however, has little
11 evolutionary significance unless retained haplotypes subsequently contribute to adaptation.
12 Our analyses provide evidence for such a transition in the centromeric region of PaChr7A,
13 where a large retained haploblock shows strong signatures of positive selection across
14 independently evolved saline populations. The concordance between selective sweep signals,
15 population differentiation, transcriptomic divergence, and functional enrichment indicates that
16 this region has been repeatedly recruited during adaptation to saline environments, rather than
17 simply persisting as a selectively neutral remnant. A comparable pattern has been reported in
18 the large yellow croaker (*Larimichthys crocea*), where centromeric regions exhibit suppressed
19 recombination, reduced genetic diversity, and elevated population differentiation together with
20 signatures of positive selection, indicating that centromeres can also harbor adaptive genetic
21 variation⁴¹. These examples show how centromeric ancestry can move beyond passive
22 persistence to acquire ecological significance, with suppressed recombination preserving
23 adaptive allele combinations that are later favored under environmental selection. While some
24 centromeric haploblocks likely represent ancestral components retained during establishment
25 of the octoploid lineage, others show clear signatures of subsequent cross-ploidy introgression,
26 including repeated exchange between the AU and EU lineages. By contrast, the inversion
27 identified on PaChr22A appears to have restricted ancestry exchange despite suppressed
28 recombination, showing that local chromosome architecture can either facilitate or impede the
29 long-term persistence of cross-ploidy ancestry. Nevertheless, inversions are also well known
30 to promote the divergence and maintenance of adaptive loci under ongoing gene flow⁴². A
31 parallel case on PaChr4A, where the EU and Med lineages harbor a centromere divergent from
32 both the CN tetraploid and AU octoploid lineages, suggests that potential inversion-mediated
33 centromeric isolation may be a recurrent mechanism generating centromere novelty within the
34 complex. A mechanistic precedent for this comes from the Atlantic silverside, where inversions

1 overlapping centromeric regions similarly suppress recombination and accelerate local genetic
2 differentiation⁴². Likewise, chromosomal inversions have repeatedly been associated with
3 adaptive phenotypic divergence in *Drosophila* and the Atlantic silverside by maintaining co-
4 adapted haplotypes under environmentally driven divergent selection⁴³⁻⁴⁵. Together, these
5 cases reframe the centromere paradox: the same recombination suppression that allows
6 centromeric regions to serve as conduits for adaptive introgression can, when combined with
7 structural rearrangements, instead accelerate centromeric divergence and reinforce
8 reproductive isolation^{39,46}. Whether a given centromeric region becomes a hotspot of exchange
9 or a barrier to gene flow thus appears to depend on local structural context.

10 More broadly, large haploblocks underlying parallel adaptation have been documented in
11 other plant systems, including *Trifolium repens*⁴⁷ and *Ambrosia artemisiifolia*⁴⁸, but the present
12 case is distinctive in that the haploblock is pericentromeric in origin and was acquired through
13 cross-ploidy rather than within-ploidy gene flow. Given the pervasiveness of polyploidy and
14 recurrent hybridization across the Poaceae, this centromere-mediated introgression framework
15 may illuminate a general mechanism underlying the repeated ecological success of polyploid
16 grasses in environmentally heterogeneous landscapes, with potential relevance to stress
17 tolerance in polyploid crop species. Rather than functioning solely as chromosome-segregation
18 domains, centromeres may also serve as evolutionary reservoirs that preserve and recycle
19 adaptive genetic variation during polyploid evolution, however, whether this mechanism is
20 widespread remains to be tested.

23 **Materials and Methods**

24 **Whole genome sequencing and preprocessing**

25 Raw reads were quality controlled using Trimmomatic⁴⁹, with adapter sequences removed and
26 low-quality bases trimmed using the parameters ILLUMINACLIP:TruSeq3-PE.fa:2:30:10
27 SLIDINGWINDOW:4:15 MINLEN:50. The resulting clean paired-end reads were aligned to
28 the reference genome using BWA-MEM⁵⁰ and coordinate-sorted with SAMtools⁵¹. PCR
29 duplicates were subsequently marked and removed using GATK MarkDuplicates⁵². Reads with
30 mapping quality scores below 30 and those with supplementary alignments were filtered out
31 using SAMtools with the flags -q 30 -F 2048. Ploidy levels were assigned based on prior studies,
32 determined either by flow cytometry or inferred from alternative allele frequency distributions
33 following genome mapping³². Individuals belonging to the AU lineage are octoploid, whereas
34 those in the USnat, EU, Med, and CN lineages are tetraploid. The reference genome of *P.*

australis was obtained from our previous work (NCBI accession: ASM4037322v1)⁵³. SNP calling was performed per-sample using GATK HaplotypeCaller in GVCF mode, followed by joint genotyping with GATK GenotypeGVCFs across all 25 chromosomes, with ploidy levels specified accordingly. The resulting raw SNPs were hard-filtered using GATK VariantFiltration with the following thresholds: MQ < 30.0, FS > 60.0, SOR > 3.0, and ExcessHet > 20.0. To reduce the effects of linkage disequilibrium, SNPs were subsequently thinned using GATK SelectVariants with a window size of 10 bp and a step size of 10,000 bp, retaining one representative SNP per window.

Phylogeny and population genetics

To avoid confounding effects of organellar DNA, only SNPs mapped to chromosomes 1–25 were retained, excluding mitochondrial and chloroplast sequences. The thinned VCF files were converted to PHYLIP format using the script `vcf2phylip.py` (<https://github.com/edgardomortiz/vcf2phylip>). A phylogenetic tree was inferred for each chromosome independently using RAxML-NG⁵⁴ under the GTR+G substitution model with 100 bootstrap replicates. The resulting per-chromosome trees were then used as input for ASTRAL⁵⁵ to estimate a coalescent-based species tree. Both the per-chromosome trees and the species tree were visualized using iTOL⁵⁶, with only nodes receiving bootstrap support values ≥ 0.8 displayed. The thinned VCF files were concatenated using BCFtools concat to generate a whole-genome SNP dataset, yielding a total of 464,630 SNPs. Principal component analysis (PCA) was performed using SNPRelate⁵⁷, with alternative allele dosage counts relative to the reference genome used as variant input to appropriately handle polyploid genotypes.

To infer patterns of genetic admixture among populations, clustering analyses were performed using ADMIXTURE⁵⁸. SNPs were further filtered to remove linkage disequilibrium using PLINK⁵⁹ with the parameter `--indep-pairwise 50 10 0.2`. Three datasets were analyzed separately: (1) all species combined and (2) tetraploid *P. australis* populations only. For each dataset, values of K ranging from 2 to 12 were evaluated, with 10 replicates per K to obtain cross-validation (CV) error estimates. The K value yielding the lowest CV error was selected as the optimal K; however, multiple K values were also visualized to provide a more comprehensive interpretation of population structure. To reconstruct the demographic history of each tetraploid lineage, changes in effective population size (N_e) over time were inferred using SMC++⁶⁰. Eight samples were selected for each of the CN tetraploid, *P. karka*, *P. mauritanus*, and EU lineages, and five samples for the Med lineage. A mutation rate of 2.17×10^{-9} per site per generation²³, was applied as the molecular clock. For each analysis, each

individual was iteratively designated as the distinguished individual across separate runs, and the final N_e trajectory was estimated by aggregating results across all runs and all 25 chromosomes. Introgression among phylogenetic groups was assessed using TWISST⁶¹. For each chromosome, individuals from the USnat population were designated as the outgroup as it represents the earliest diverging lineage within *P. australis*, with 55 AU octoploid individuals, 27 CN tetraploid individuals, and 23 EU and Med individuals included in the analysis. Per-chromosome phylogenetic trees were constructed in overlapping windows and weighted topology sorting was performed using a smoothed window size of 20 kb.

Genetic diversity, F_{ST} and genome scan

Based on the population structure analyses, genetic differentiation was detected among ecotypes of *P. australis* within the CN tetraploid lineage, specifically among sea, freshwater, and salt-lake populations. To investigate whether these ecotypes have experienced natural selection, nucleotide diversity (π and Watterson's θ) and pairwise F_{ST} were calculated following the methods outlined in Pieri *et al.*⁴² and Wang *et al.*¹¹. Briefly, ancestral alleles were first inferred by mapping publicly available FASTQ files of six Poaceae species, obtained from Wang *et al.*⁽¹¹⁾, to the *P. australis* reference genome. The ancestral state at each site was determined as the most frequently observed base across outgroup species using ANGSD⁶². Based on this ancestral information, π and Watterson's θ were estimated from the unfolded site frequency spectrum (uSFS) using individual BAM files as input. Pairwise F_{ST} between brackish water and freshwater populations within the CN tetraploid lineage was additionally estimated using ANGSD with a sliding window size of 10 kb. SNPs falling within the top 1% of the F_{ST} distribution were considered candidate loci potentially under divergent selection. To further characterize regions potentially under selection, population branch statistic (PBS) values were calculated alongside π , and candidate selective sweeps were identified by jointly visualizing PBS and π across the genome.

Centromere identification

Centromeres of the 25 chromosomes were identified using a combination of indices, including Hi-C contact maps, repeat density, and gene density. Hi-C reads for *P. australis* were obtained from NCBI (accession number SRR29533168) and mapped to the reference genome using the Juicer pipeline⁶³. Hi-C contact heatmaps were visualized along each chromosome using pyGenomeTracks⁶⁴. Tandem repeats were identified using TideCluster (<https://github.com/kavonrtep/TideCluster>) implemented in RepeatExplorer⁶⁵. Satellite DNA was identified using the TRASH pipeline⁶⁶ and subsequently classified into clusters.

Centromeric regions were defined as genomic intervals exhibiting high repeat density, low gene density, and enriched local Hi-C contacts and reduced long-range interactions with flanking chromosomal arms. These information were then combined and used to identify the centromeres via CentIER⁶⁷.

Introgression analysis using ABBA-BABA (D-statistic) tests

To test for introgression among lineages of *P. australis*, we performed ABBA-BABA (Patterson's D-statistic) tests based on derived allele frequencies. Biallelic SNPs were extracted from the filtered VCF files for each chromosome using bcftools v1.16⁵¹. Because sample ploidy varied, genotypes were parsed in a ploidy-aware manner using parseVCF.py from the genomics_general package (https://github.com/simonhmartin/genomics_general), with per-sample ploidy specified in a ploidy file. Derived allele frequencies for each population (CN, EU, AU, and the outgroup *P. japonicus*) were calculated at each site using freq.py from the same package. For the topology (((P1, P2), P3), O) = (((CN, EU), AU), PJ), the D-statistic was calculated as: $D = (\Sigma ABBA - \Sigma BABA) / (\Sigma ABBA + \Sigma BABA)$, and p1, p2, and p3 are the derived allele frequencies in P1, P2, and P3, respectively. A significant positive D indicates excess allele sharing between P2 and P3 relative to P1, consistent with introgression between the EU and AU lineages, whereas a significant negative D indicates excess sharing between P1 and P3, consistent with introgression between the CN and AU lineages⁶⁸. D-statistics were calculated both genome-wide and separately for each chromosome. Standard errors were estimated by block-jackknife resampling, in which the genome was divided into non-overlapping 1-Mb physical blocks and each block was sequentially removed to generate a distribution of jackknife pseudo-values; the standard error was calculated from the variance of these pseudo-values following Green *et al.* 2010⁶⁹. For chromosomes yielding fewer than 20 jackknife blocks at 1 Mb, block size was reduced to 200 kb to ensure reliable variance estimation. Z-scores were calculated as D divided by its standard error, and two-tailed p-values were derived from the standard normal distribution. p-values across chromosomes were adjusted for multiple testing using the Benjamini–Hochberg false discovery rate procedure⁷⁰, with $p_{adj} < 0.05$ considered statistically significant. To characterize heterogeneity in introgression signal along each chromosome, D-statistics were calculated in sliding windows using the script ABBABABAwindows.py. The same approach was applied to the topology (((CN, Med), AU), PJ) to test for potential introgression between the Med and AU lineages.

Identification of population-specific SNPs and subgenome origin of AU octoploids

To identify ecotype-specific SNPs between the marine and salt-lake tetraploid populations of *P. australis*, we analyzed variant call data from 5 marine (TS) and 7 salt-lake (YH) allotetraploid individuals. All variants were called against the chromosome-level *P. australis* reference genome. Initial filtering retained only biallelic SNPs with a per-site missing genotype rate ≤ 2 individuals out of 12, yielding 274,524 high-quality SNPs for downstream analysis. Allele frequencies were calculated for each population independently using scikit-allel (<https://zenodo.org/records/13772087>). A SNP was defined as population-specific when the alternative allele was completely fixed in one population (frequency = 1.0) and entirely absent in the other (frequency = 0.0). This stringent criterion identified 7,751 marine-specific SNPs and 24,362 salt-lake-specific SNPs. All analyses were restricted to the PaChr7A genomic windows exhibiting elevated F_{ST} . To investigate the subgenome origin of the AU octoploid, we extracted the allele frequencies of the marine- and salt-lake-specific SNPs from 55 AU octoploid individuals using the same variant call set. Given the octoploid nature of these individuals, allele frequencies were calculated as the proportion of alternative allele observations across all sequencing reads at each locus, without assuming a fixed ploidy. A population-specific allele was considered present in the AU octoploid if its frequency exceeded 0.05.

Plant maintenance and salt treatment

The common garden was maintained in Changle County, Weifang City, Shandong Province, China (36°22' N, 118°53' E). We collected 126 individuals (genotypes) of *P. australis* rhizomes from the common garden for cultivation. The populations from China were categorized into three ecotypes reflecting their source habitats: freshwater wetland, coastal saltmarsh, and salt-lake types. Ecotype classification was not applied to non-Chinese populations due to insufficient habitat documentation at the time of collection. The plants were planted in barrels (total volume 25 L; top diameter 32.5 cm, bottom diameter 28.4 cm, height 38.5 cm) filled with 20 L of a soil substrate mixed in a 2 (soil):1 (peat moss):1 (river sand) ratio. The salinity treatment was applied as a 1% NaCl solution (10 ppt), calculated based on the estimated porewater volume of the 20 L soil system. Each individual was planted in two barrels: one for the salt treatment and one for the control. This concentration falls within the range encountered by *P. australis* in saline wetlands, and represents an ecologically recurrent stress that is moderate for most lineages of *P. australis* but severe for specific genotypes (e.g., those native to North America)^{21,71}, thus capturing the natural variation in salinity tolerance across its global distribution. This was achieved by adding 100 g of NaCl dissolved in water to each barrel on

June 2 and 9, for a total of 200 g. Green leaves were sampled at one month after salt addition (July 10, 2024) and used for transcriptomic analyses.

RNAseq library preparation

The plant total RNA was extracted using the RNAPrep Pure Plant Kit (Tiangen, Beijing, China) according to the instructions provided by the manufacturer. RNA concentration and purity was measured using NanoDrop 2000(Thermo Fisher Scientific, Wilmington, DE). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). A total amount of 1 µg RNA per sample was used as input material for the preparations of RNA samples. Sequencing libraries were generated using Hieff NGS Ultima Dual-mode mRNA Library Prep Kit for Illumina (Yeasten Biotechnology (Shanghai) Co., Ltd.) following manufacturer's recommendations and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. First strand cDNA was synthesized, and second strand cDNA synthesis was subsequently performed. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEBNext Adaptors with hairpin loop structure were ligated to prepare for hybridization. The library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, USA). Then 3 µl USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95°C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system. The libraries were sequenced on an Illumina NovaSeq 6000 platform to generate 150 bp paired-end reads, according to the manufacturer's instructions.

Read preprocessing and DEG analyses

Quality of the reads were checked using fastqc (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adapters were clipped and low quality bases (phred score < 20) were removed from the reads in 4 bp sliding windows using Trimmomatic⁴⁹. Reads and bases with a Phred score higher than 20 were retained⁷². For transcriptomic data analysis, clean reads were aligned to the *P. australis* reference genome of a Chinese individual, using the STAR aligner two-pass procedure⁷³. After sorting using Samtools v1.16.1⁵¹, the mapped reads were processed using StringTie⁷⁴ to obtain the gene and transcript count matrix. Principal component analysis (PCA) was conducted on differentially

expressed genes among the treatment groups using DESeq2 following rlog transformation of gene counts, which represent the number of reads mapped to the reference genome ⁷⁵.

DESeq2 provides statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted P-value < 0.01 & Fold Change ≥ 2 found by DESeq2 were assigned as differentially expressed. Gene Ontology (GO) enrichment assessment analyses sets of genes in the context of GO categories to determine whether certain functional categories are overrepresented or enriched within the background whole genome gene sets. Here we tested enriched GO terms for the differentially expressed genes (DEGs) using goatools ⁷⁶, using a criterion of P < 0.05 with Bonferroni correction. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was performed using KAAS ⁷⁷ and the enriched pathway of differentially expressed genes were analyzed using clusterProfiler ⁷⁸. The Venn diagram was performed using ggvenn ⁷⁹.

WGCNA

Module eigengenes were calculated with moduleEigengenes and ordered with orderMEs. Module-trait relationships were assessed by Pearson correlation between module eigengenes and a one-hot-encoded condition matrix, with p-values from corPvalueStudent. Hub genes for the module of interest were identified by intramodular connectivity, computed as row sums of an unsigned adjacency matrix (adjacency, power = 9, type = "unsigned") restricted to genes assigned to that module, ranked in descending order.

Co-expression networks were built with the WGCNA package in R⁸⁰. Raw counts were normalized in DESeq2 using the variance-stabilizing transformation (vst, blind = TRUE) across all samples, yielding a genome-wide expression matrix (41,284 genes retained). The soft-thresholding power was evaluated with pickSoftThreshold under signed and unsigned settings (powers 1–20), and a power of 9 was selected by visual inspection of the scale-free topology fit rather than by an automatic estimate. A signed network was constructed with blockwiseModules (power = 9, networkType = "signed", TOMType = "signed", minModuleSize = 50, mergeCutHeight = 0.30, pamRespectsDendro = FALSE, maxBlockSize set above the total gene number), and module labels were converted to colors with labels2colors. Module eigengenes were calculated with moduleEigengenes and ordered with orderMEs; the grey (unassigned) module was excluded from all downstream eigengene and trait-correlation analyses. Module-trait relationships were assessed by Pearson correlation between module eigengenes and a one-hot-encoded condition matrix (model.matrix(~0+condition)); p-values

were computed with `corPvalueStudent` and corrected for multiple testing across the full module-by-trait matrix using the Benjamini–Hochberg method. For each condition, modules with a significant positive association ($\text{FDR} < 0.05$, $r > 0$) were identified computationally; a module with a notable negative association (darkred with AU) was additionally identified by visual inspection of the heatmap. Module membership (kME) was computed with `signedKME` on the grey-excluded eigengene set. For the modules of interest (greenyellow and darkred, associated with AU; lightyellow, associated with EUinv; black, associated with Land; and lightgreen, examined separately), genes were ranked by descending kME within each module and used in full (without a fixed top-N cutoff) as the module gene set. GO enrichment was performed with GOATOOLS (`find_enrichment`) for each module's full gene list against a genome-wide population file, at a nominal enrichment threshold of $P < 0.05$; for the greenyellow module, an additional enrichment test restricted to the top 130 kME-ranked genes was performed as a supplementary analysis.

Data availability

The whole genome sequencing and RNAseq data for the 226 samples were deposited in NCBI SRA database, with the project accession number PRJNA1507497.

Acknowledgments

The authors thank Miss Dongzi Jiang for her contribution to the greenhouse experiment. This work is supported by the National Natural Science Foundation of China (No. 32100304; 32470388; U22A20558; 32271588) and the Natural Science Foundation of Shandong Province (ZR2021QC119). We thank CSC–IT Center for Science, Finland for computation resources.

Author contributions:

LL. Liu and WH. Guo conceived the idea; LL. Liu, WY. Sheng, HJ. Song, YH. Wang maintained the plant and carried out the salt treatment experiments; LL. Liu and WY. Sheng generated the genomic and transcriptomic data; C. Wang performed the genomic and transcriptomic data analyses with contributions from WB. Luo and M. Ilievska; C. W. led the writing of the paper with contribution from LL. Liu, WH. Guo, WY. Sheng, WB. Luo, LL. Lin, M. Ilievska, HJ. Song, and YH. Wang.

References

1. Figueiredo, M. *et al.* Intraspecific and interspecific polyploidy of Brazilian species of the genus *Inga* (Leguminosae: Mimosoideae). *Genetics and Molecular Research* **13**, 3395–3403 (2014).
2. Wagner, F. *et al.* ‘At the crossroads towards polyploidy’: genomic divergence and extent of homoploid hybridization are drivers for the formation of the ox-eye daisy polyploid complex (*Leucanthemum*, Compositae-Anthemideae). *New Phytologist* **223**, 2039–2053 (2019).
3. Brysting, A.K., Fay, M.F., Leitch, I.J. & Aiken, S.G. One or more species in the arctic grass genus *Dupontia*?—a contribution to the Panarctic Flora project. *Taxon* **53**, 365–382 (2004).
4. Cordeiro, J.M.P. & Felix, L.P. Intra-and interspecific karyotypic variations of the genus *Senna* Mill. (Fabaceae, Caesalpinioideae). *Acta Botanica Brasilica* **32**, 128–134 (2018).
5. Del Pozo, J.C. & Ramirez-Parra, E. Whole genome duplications in plants: an overview from *Arabidopsis*. *Journal of Experimental Botany* **66**, 6991–7003 (2015).
6. Coexistence, G.-F.S.C. Frequent Origins and Unidirectional Gene-Flow Sustain Cytotype Coexistence in a Mixed-Ploidy Species.
7. Baduel, P., Hunter, B., Yeola, S. & Bomblies, K. Genetic basis and evolution of rapid cycling in railway populations of tetraploid *Arabidopsis arenosa*. *PLoS genetics* **14**, e1007510 (2018).
8. Henikoff, S., Ahmad, K. & Malik, H.S. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**, 1098–1102 (2001).
9. Yang, X. *et al.* Amplification and adaptation of centromeric repeats in polyploid switchgrass species. *New Phytologist* **218**, 1645–1657 (2018).
10. Yadav, V., Sun, S., Coelho, M.A. & Heitman, J. Centromere scission drives chromosome shuffling and reproductive isolation. *Proceedings of the National Academy of Sciences* **117**, 7917–7928 (2020).
11. Wilkinson, M.J. *et al.* Centromeres are hotspots for chromosomal inversions and breeding traits in mango. *New Phytologist* **245**, 899–913 (2025).
12. Gerchen, J.F., Mandák, B., Melnyk, M. & Kolář, F. Centromeric variation is shared across ploidy barriers in *Alnus glutinosa* agg. *bioRxiv*, 2026.04. 21.719804 (2026).
13. Wang, C. *et al.* Genome-wide analysis tracks the emergence of intraspecific polyploids in *Phragmites australis*. *npj Biodiversity* **3**, 29 (2024).
14. Cui, S. *et al.* Proteomic characterization of *Phragmites communis* in ecotypes of swamp and desert dune. *Proteomics* **9**, 3950–3967 (2009).
15. Qiu, T. & Cui, S. Evolutionary analysis for *Phragmites* ecotypes based on full-length plastomes. *Aquatic Botany* **170**, 103349 (2021).
16. Lambertini, C. *et al.* A phylogeographic study of the cosmopolitan genus *Phragmites* (Poaceae) based on AFLPs. *Plant Systematics and Evolution* **258**, 161–182 (2006).
17. Saltonstall, K. Cryptic invasion by a non-native genotype of the common reed, *Phragmites australis*, into North America. *Proceedings of the National Academy of Sciences* **99**, 2445–2449 (2002).
18. Wang, C. *et al.* Genome-wide analysis tracks the emergence of intraspecific polyploids in *Phragmites australis*. *bioRxiv*, 2021.09. 05.458733 (2021).
19. Liu, L.L. *et al.* Cryptic lineages and potential introgression in a mixed-ploidy species (*Phragmites australis*) across temperate China. *Journal of Systematics and Evolution* **60**, 398–410 (2022).

20. Saltonstall, K. The naming of *Phragmites* haplotypes. *Biological Invasions* **18**, 2433–2441 (2016).
21. Eller, F. *et al.* Cosmopolitan species as models for ecophysiological responses to global change: the common reed *Phragmites australis*. *Frontiers in plant science* **8**, 1833 (2017).
22. Wang, C. *et al.* Single-cell and spatial transcriptomics in *Phragmites australis* reveal the association of B chromosomes with plant invasiveness. *Genome Biology* (2026).
23. Wang, C. *et al.* Transcriptome analysis of tetraploid and octoploid common reed (*Phragmites australis*). *Frontiers in Plant Science* **12**, 653183 (2021).
24. Srivastava, J., Kalra, S.J. & Naraian, R. Environmental perspectives of *Phragmites australis* (Cav.) Trin. Ex. Steudel. *Applied Water Science* **4**, 193–202 (2014).
25. Wang, C. *et al.* Chromosome-level genome assemblies reveal genome evolution of an invasive plant *Phragmites australis*. *Communications Biology* **7**, 1007 (2024).
26. Tanaka, T.S., Irbis, C. & Inamura, T. Phylogenetic analyses of *Phragmites* spp. in southwest China identified two lineages and their hybrids. *Plant Systematics and Evolution* **303**, 699–707 (2017).
27. Lambertini, C. *et al.* Phylogenetic diversity shapes salt tolerance in *Phragmites australis* estuarine populations in East China. *Scientific reports* **10**, 17645 (2020).
28. Meyerson, L.A. *et al.* Do ploidy level and nuclear genome size and latitude of origin modify the expression of *Phragmites australis* traits and interactions with herbivores? *Biological Invasions* **18**, 2531–2549 (2016).
29. Te Beest, M. *et al.* The more the better? The role of polyploidy in facilitating plant invasions. *Annals of botany* **109**, 19–45 (2012).
30. Guan, B. *et al.* Clonal integration promotes the growth of *Phragmites australis* populations in saline wetlands of the Yellow River Delta. *Frontiers in Plant Science* **14**, 1162923 (2023).
31. Sun, X.-S. *et al.* Effects of salinity and concomitant species on growth of *Phragmites australis* populations at different levels of genetic diversity. *Science of The Total Environment* **780**, 146516 (2021).
32. Wang, C. *et al.* Genome-wide analysis illustrates the effect of polyploidization in *Phragmites australis*. *bioRxiv* (2021).
33. Lambertini, C. *et al.* A phylogeographic study of the cosmopolitan genus *Phragmites* (Poaceae) based on AFLPs. *Plant Systematics and Evolution* **258**, 161–182 (2006).
34. Yang, Z. *et al.* Cotton D genome assemblies built with long-read data unveil mechanisms of centromere evolution and stress tolerance divergence. *BMC biology* **19**, 115 (2021).
35. Cho, M. *et al.* Block of ATP-binding cassette B19 ion channel activity by 5-nitro-2-(3-phenylpropylamino)-benzoic acid impairs polar auxin transport and root gravitropism. *Plant physiology* **166**, 2091–2099 (2014).
36. Brown, M.R., Abbott, R.J. & Twyford, A.D. The emerging importance of cross-ploidy hybridisation and introgression. *Molecular Ecology* **33**, e17315 (2024).
37. Wang, T. *et al.* Interploidy introgression shaped adaptation during the origin and domestication history of *Brassica napus*. *Molecular Biology and Evolution* **40**, msad199 (2023).
38. Han, J. *et al.* Centromere-size reduction and chromatin state dynamics following intergenomic hybridization in cotton. *PLoS genetics* **21**, e1011689 (2025).

39. Zhao, J. *et al.* Centromere repositioning and shifts in wheat evolution. *Plant Communications* **4**(2023).
40. Gao, S. *et al.* A global view of human centromere variation and evolution. *Nature*, 1–12 (2026).
41. Cui, Y. *et al.* T2T genomes unveil centromere architecture and adaptive divergence in large yellow croaker (*Larimichthys crocea*). *Advanced Science* **12**, e06374 (2025).
42. Akopyan, M., Tigano, A., Jacobs, A., Wilder, A.P. & Therkildsen, N.O. Genetic differentiation is constrained to chromosomal inversions and putative centromeres in locally adapted populations with higher gene flow. *Molecular Biology and Evolution* **42**, msaf092 (2025).
43. Lenhart, B.A. & Bergland, A.O. Cosmopolitan inversions have a major impact on trait variation and the power of different GWAS approaches to identify associations. *PLoS Genetics* **22**, e1012012 (2026).
44. Akopyan, M. *et al.* Multiple chromosomal inversions modulate continuous local adaptation along a steep thermal cline. *Science* **391**, 1015–1021 (2026).
45. Poikela, N., Hoikkala, V., Ritchie, M.G. & Kankare, M. Examining the role of a chromosomal inversion in accumulating adaptive and barrier loci in a cold-adapted insect species. *Genome Biology and Evolution* **18**, evag101 (2026).
46. Santaguida, S. & Musacchio, A. The life and miracles of kinetochores. *The EMBO journal* **28**, 2511 (2009).
47. Battlay, P. *et al.* Haploblocks contribute to parallel climate adaptation following global invasion of a cosmopolitan plant. *Nature ecology & evolution* **9**, 1441–1455 (2025).
48. Battlay, P. *et al.* Rapid parallel adaptation in distinct invasions of *Ambrosia Artemisiifolia* is driven by large-effect structural variants. *Molecular Biology and Evolution* **42**, msae270 (2025).
49. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
50. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25**, 1754–1760 (2009).
51. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
52. Van der Auwera, G.A. & O'Connor, B.D. *Genomics in the cloud: using Docker, GATK, and WDL in Terra*, (O'Reilly Media, 2020).
53. Wang, C. *et al.* Chromosome-level genome assemblies reveal genome evolution of an invasive plant *Phragmites australis*. *Communications Biology* **7**, 1007 (2024).
54. Kozlov, A.M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
55. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics* **19**, 153 (2018).
56. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic acids research* **52**, W78–W82 (2024).
57. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).

58. Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* **19**, 1655–1664 (2009).
59. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics* **81**, 559–575 (2007).
60. Terhorst, J., Kamm, J.A. & Song, Y.S. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature genetics* **49**, 303–309 (2017).
61. Martin, S.H. & Van Belleghem, S.M. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429–438 (2017).
62. Korneliussen, T.S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC bioinformatics* **15**, 1–13 (2014).
63. Durand, N.C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell systems* **3**, 95–98 (2016).
64. Ramírez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature communications* **9**, 189 (2018).
65. Novák, P., Neumann, P., Pech, J., Steinhaisl, J. & Macas, J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **29**, 792–793 (2013).
66. Wlodzimierz, P., Hong, M. & Henderson, I.R. TRASH: tandem repeat annotation and structural hierarchy. *Bioinformatics* **39**, btad308 (2023).
67. Xu, D. *et al.* CentIER: Accurate centromere identification for plant genomes. *Plant communications* **5**(2024).
68. Martin, S.H., Davey, J.W. & Jiggins, C.D. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Molecular biology and evolution* **32**, 244–257 (2015).
69. Green, R.E. *et al.* A draft sequence of the Neandertal genome. *science* **328**, 710–722 (2010).
70. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289–300 (1995).
71. Meyerson, L.A., Cronin, J.T., Packer, J., Pyšek, P. & Saltonstall, K. Ecology, Evolution, and North American Invasion of One of the World's Most Successful Plant Species. *Annual Review of Ecology, Evolution, and Systematics* **56**, 73–98 (2025).
72. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10–12 (2011).
73. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
74. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology* **33**, 290–295 (2015).
75. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 1–21 (2014).
76. Klopfenstein, D. *et al.* GOATOOLS: A Python library for Gene Ontology analyses. *Scientific reports* **8**, 10872 (2018).
77. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic acids research* **35**, W182–W185 (2007).
78. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The innovation* **2**(2021).

- 1 79. Yan, L. & Yan, M.L. Package 'ggvenn'. (Google Scholar, 2021).
- 2 80. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network
- 3 analysis. *BMC bioinformatics* **9**, 559 (2008).
- 4

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplemenataryNotes.pdf](#)
- [SupplementaryMaterials.pdf](#)
- [Tables.xlsx](#)