Single-cell and spatial transcriptomics uncover the role of B chromosomes in driving plant invasiveness

Cui Wang[1,2], James Ord[1], Mengxiao Yan[3], Hongjin Shao[2], Lele Lin[4], Jarkko Salojärvi[1,5], Lele Liu[2]*, Weihua Guo[2]*

Authors: Cui Wang, Dr., 1 Organismal and Evolutionary Biology Research Program, Faculty of Biological and Environmental Sciences, University of Helsinki, Viikinkaari 1, Biocentre 3, 00790 Helsinki, Finland

2 Qingdao Key Laboratory of Ecological Protection and Restoration, School of Life Sciences, Shandong University, 72 Binhai Road, Qingdao 266237, China

James Ord, Dr., 1 Organismal and Evolutionary Biology Research Program, Faculty of Biological and Environmental Sciences, University of Helsinki, Viikinkaari 1, Biocentre 3, 00790 Helsinki, Finland

Mengxiao Yan, Dr., 3 Shanghai Key Laboratory of Plant Functional Genomics and Resources, Shanghai Chenshan Botanical Garden, 201602, Shanghai, China

Hongjin Shao, Mr., 2 Qingdao Key Laboratory of Ecological Protection and Restoration, School of Life Sciences, Shandong University, 72 Binhai Road, Qingdao 266237, China

Lele Lin, Dr., 4 Ecology and Nature Conservation Institute, Chinese Academy of Forestry, No.1 Dongxiaofu, Xiangshan Road, Haidian district, Beijing 100091, China

Jarkko Salojärvi, Assistant Professor, 1 Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Viikinkaari 1, Biocentre 3, 00790 Helsinki, Finland

5 School of Biological Sciences, Nanyang Technological University, Singapore

Lele Liu*, Dr., 2 Qingdao Key Laboratory of Ecological Protection and Restoration, School of Life Sciences, Shandong University, 72 Binhai Road, Qingdao 266237, China

Weihua Guo*, Professor., 2 Qingdao Key Laboratory of Ecological Protection and Restoration, School of Life Sciences, Shandong University, 72 Binhai Road, Qingdao 266237, China

Correspondence to:

Dr. Lele Liu, liulele@sdu.edu.cn

Professor Weihua Guo, whguo@sdu.edu.cn

Abstract
Invasive plants can profoundly disrupt native biodiversity, yet the genetic mechanisms underpinning their success remain poorly understood. To date, genomic studies have been conducted on only a limited number of invasive species, and no single-cell level studies have been applied. This research investigates the genetic drivers behind the invasive behavior of the common reed (*Phragmites australis*), a hardy grass species that became invasive in North America following its introduction from Europe. By integrating whole-genome sequencing with single-cell and spatial transcriptomics, we developed a comprehensive single-cell atlas of the common reed's shoot system. UMAP analysis identified 19 distinct cell clusters within the shoot system. Gene Ontology (GO) enrichment enabled the annotation of key cell types, including mesophyll, epidermal, bundle sheath, and xylem cells, as well as shoot apical and lateral bud meristems, and auxillary meristems. RNA velocity analysis highlighted the multipotent nature of mesophyll cells, with chlorenchyma in Cluster 3 identified as progenitor cells capable of differentiating into various tissues and Cluster 1 progressing towards aerenchyma formation. Comparative analysis between European and North American invasive populations revealed significant differences in transcriptional activity and gene expression, particularly in clusters associated with the shoot apical meristem. Invasive populations exhibited a higher prevalence of B chromosomes, with three genes IMPA-3, SSC3, and DDE family endonuclease gene consistently upregulated across nearly all clusters, notably near mesophyll and meristematic regions. The fast mutation in IMPA-3 which functions as a major receptor of Resistance (R) gene may have strengthened the adaptability of the invasive population in North America. These findings provide critical insights into the cellular development and genomic diversity underlying the invasiveness of the common reed, offering valuable information to guide ecological management strategies.

Key words: Single cell transcriptomics, Spatial transcriptomics, Invasive plants, B chromosome

Introduction

Bioinvasion is considered one of the most significant threats to ecosystems worldwide, as non-native species can disrupt ecological balance, outcompete native species, and alter habitat structure(*1*). While many species become invasive upon introduction to new environments, some species may remain non-invasive in their native habitats but exhibit invasive characteristics when introduced to novel ecosystems. For example, species such as *Spartina* spp., *Ambrosia artemisiifolia* L., and *Phragmites australis* are non-invasive in their native regions, yet become highly invasive in new environments *(2, 3)*. Invasive populations often exhibit distinct phenotypic or reproductive traits in the region they have invaded, which may differ significantly from those observed in their native range. These contrasting traits, such as altered growth rates, reproductive strategies, or resource use, can play a pivotal role in their invasive success(*4, 5*). Individuals displaying invasive characteristics, such as enhanced competitive abilities or adaptive responses to the new environment may arise in the introduced population, even if these traits were not present in the original population.

To better understand the invasion process, it is crucial to study populations from both their native and introduced ranges. By comparing these populations, we can identify the key traits and genetic variations that distinguish invasive individuals from their native counterparts. Various genetic factors have been implicated in facilitating bioinvasion. These include additive genetic variance, where multiple genes contribute incrementally to an individual's fitness in the new environment; epistasis, which refers to interactions between genes that influence phenotypic traits; hybridization, where gene flow between different populations or species can introduce new traits; genetic trade-offs, where adaptations to one ecological factor may come

at a cost to others; and the action of a small number of genes that might have a disproportionately large effect on invasiveness. Genomic rearrangements may also play a role in enabling populations to rapidly adapt to novel conditions. Understanding these genetic mechanisms is essential for uncovering how species become invasive and how they can outcompete native species in newly occupied habitats(*6*). Traditional methods for studying the processes and mechanisms of bioinvasions have typically relied on large-scale ecological observations or phylogenetic studies that use a limited number of genetic markers to infer the origins and dynamics of invasive species. With the growing accessibility of sequencing technologies, next-generation sequencing (NGS) has emerged as a powerful tool in bioinvasion research. In recent years, the use of advanced technologies, such as single-cell and spatial transcriptomics, has expanded significantly. However, most of these studies remain focused on *Arabidopsis*, leaving non-model species largely understudied. Genomic data have shown that the impact of multiple introductions into new environments can surpass the influence of genetic diversity in shaping invasive populations, often giving them a competitive edge over local populations (*7*).

The common reed (*Phragmites australis*) is one of the most globally widespread plant species, thriving in a variety of wetland habitats across temperate and tropical regions. Its extensive distribution is accompanied by remarkable intraspecific genetic diversity, with significant variation in genetic loci, ploidy levels, and morphological traits (*8-10*). This diversity has allowed *P. australis* to adapt to a wide range of environmental conditions, contributing to its ecological resilience and competitive success. Despite its native status in Europe, where it plays an essential role in wetland ecosystems without exhibiting invasive tendencies, *P. australis* has become a highly aggressive invasive species in North America. Its introduction to the continent is believed to have occurred approximately 200 years ago(*3*). Since then, *P. australis* has expanded rapidly across North American wetlands, where it outcompetes native vegetation, reduces biodiversity, and alters hydrological and nutrient cycling processes(*11*).

One of the key differences observed between invasive and native populations of *P. australis* lies in their regenerative capabilities. In regeneration experiments, invasive populations demonstrated a significantly higher culm regeneration rate compared to source populations from Europe. In some cases, the invasive populations also exhibited greater rhizome regeneration, further enhancing their ability to colonize new areas and recover from environmental disturbances (*12*). This enhanced regenerative capacity provides invasive populations with a competitive edge, allowing them to establish dense monocultures that displace native species and transform ecosystem dynamics. Genomic studies have shed light on another intriguing aspect of *P. australis* invasion, that North American populations of *P. australis* tend to have slightly larger genomes than their European counterparts (*13*) (*14*). While the precise reasons for the increased genome size remain unclear, emerging evidence suggests that the accumulation of B chromosomes may play a central role. Our previous research identified that invasive populations of *P. australis* possess a higher number of B chromosomes compared to ancestral populations in Europe (*15*). B chromosomes are supernumerary chromosomes that exist alongside the standard set of A chromosomes, often without being essential for survival. However, they can influence various genetic and phenotypic traits, contributing to increased adaptability and competitive advantages in invasive populations. B chromosomes are known to carry large amounts of repetitive elements and non-coding ribosomal DNA (rDNA), and although they were traditionally thought to be genetically inert, recent studies suggest that they may actively participate in gene regulation(*16*). However, the functional role of B chromosomes in *P. australis* populations and the mechanisms by which they interact with the broader genome are still poorly understood.

In this study, we aim to investigate how the shoot system generated from rhizomes contributes to the invasion success of *P. australis* in invasive populations. To address this, we first compared the growth rates of rhizome shoots between invasive populations in North America and ancestral populations from Europe. This comparison will help us determine if differences in growth rate contribute to the success of invasive populations. Next, we sampled the newly formed shoot tissues regenerated from rhizome to perform single-cell and spatial transcriptomic analyses. By analyzing gene expression patterns at both the cellular and tissue levels, we aim to identify genes that are involved in the developmental process and compare their expression between invasive and ancestral populations. Of particular interest are genes regulated by B chromosomes, which have been implicated in modulating gene expression and contributing to invasiveness in previous studies. Thirdly, we aim to investigate the copy number variation of B chromosomes using whole genome sequencing data, and the evolution of invasiveness related genes. This allows us to explore the genomic differences, particularly those associated with B chromosomes, that may underlie their invasive potential. By identifying key genes involved in rhizome regeneration and their regulation by B chromosomes, we aim to gain deeper insights into the genetic and epigenetic mechanisms driving the invasion success of *P. australis* in North America.

Results

Mapping of the SC-RNAseq data

To establish a comprehensive single-cell atlas for the shoot system of common reed, we collected between 5 and 22 buds from each individual, pooling them into a petri dish to obtain the cell suspension for each sample (**Fig. 1A**). The cell viability ranged from 71% to 85%. After mapping the reads to the reference genome, the proportion of confidently mapped reads to the genome ranged from 75.9% to 87.4%, while the proportion of reads confidently mapped to the transcriptome ranged from 54.5% to 65.5% (**Table 1**). The number of captured cells across samples ranged from 7,154 to 15,297. The EU620 sample exhibited the highest median gene count, with 3,614 genes per cell, while the NAint61 sample had a median of 2,690 genes per cell. The mean number of reads per cell ranged from 24,167 to 59,994 (**Table 1**). The rate of doublets varied between 4% and 25%. After removing doublets and cells exhibiting high levels of mitochondrial and chloroplast expression, a total of 57,217 cells were retained for further analysis.

Table 1 Sample information and sequencing quality for the single cell and transcriptomics dataset

Sample name	Genetic group	Experiment type	Reads mapped to the genome (%)	Reads confidently mapped to the transcriptome (%)	Number of cells	Doublets removal (%)	Median gene per cell	Mean reads per cell
EU60	EU	Single cell sequencing	86.1	63.6	11,705	4	3142	30,786
EU78	EU	Single cell sequencing	87.4	65.5	12,225	4	3210	29,401
EU620	EU	Single cell sequencing	75.9	54.5	7154	6	3614	59,994

NAint61	Invasive	Single cell sequencing	83.7	61.4	8645	14	2690	42,478
NAint113	Invasive	Single cell sequencing	86.6	62.3	15,297	11	2461	24,167
NAint191	Invasive	Single cell sequencing	76.4	54.9	12,168	25	3236	33,900
NAint61	Invasive	Spatial transcriptomic (level 7)	95.12	88.07	10,867 (subspots)		1736 (per subspot)	-

Data quality of spatial transcriptomics

Buds from NAint61 were selected, fixed with OCT embedding, and cross-sectioned for spatial transcriptomics analysis (**Fig. 1B**). The spatial transcriptomic data from the invasive individual NAint61 yielded exceptionally high genome (95.12%) and transcriptome (88.07%) mapping rates after aligning the reads to the reference genome, with a sequencing saturation of 87.56%. The BMKMANU S3000 spatial transcriptomic platform enables multilevel resolution analysis, allowing us to obtain 9 levels of data by combining different numbers of spot points into one subspot, which serves as a unit for analysis. This approach facilitated in-depth exploration at various resolutions. The highest resolution, subspot level 1, with only one spot per unit, achieves subcellular resolution down to 5 µm. The number of subspots, median UMI counts per subspot, median genes per subspot, and the total number of genes detected for each subspot are summarized in **Table S1**. In the single-cell dataset, the median number of genes per cell ranged between 2,461 and 3,614. We selected level 7 for downstream analysis, which showed a median of 1,736 genes per subspot to represent cell-level expression data. For comparison, we also obtained a cell-split dataset using the BMKMANU S3000 BSTMatrix software to segment cells based on high-resolution images. This cell-split approach resulted in 78,448 cells, with a median of 332 genes per cell and a median UMI count of 372. The highest levels of gene expression were detected in young tissues, such as meristematic regions and parenchyma (**Supplementary Figure S1**).

Single cell clustering

After processing the data with CellRanger, we obtained the data matrix, which was subsequently analyzed using Seurat software to reduce the dimensionality and form cell clusters. To correct for batch effects among the samples, we applied the Harmony method (*17*) . A series of resolution values between 0 and 1 were tested, and the SC3 stability index indicated that the most stable clustering occurred at a resolution of 0.2 (**Supplementary Figures S2** and **S3**). Using a resolution of 0.8, 19 distinct clusters were defined in the Uniform Manifold Approximation and Projection (UMAP) plot and t-SNE (**Fig. 1C**, **Supplementary Figure S4**). We utilized the FindMarkers function in Seurat to identify the top 5 markers for each cluster. A heatmap of these markers showed that most markers were highly expressed and specific to their respective clusters (**Fig. 1D-E**). The cell counts for each cluster across the six *Phragmites australis* samples from European (EU) and North American invasive (NAint) populations are summarized in **Table S2** (also see **Fig. 1F**). Overall, Cluster 0 consistently contained the highest number of cells in both EU and NAint samples, indicating it as the dominant cell type

across populations. Smaller clusters, such as clusters 5, 6, 7, and 8, were also present in all samples. The markers identified for each cluster are listed in **Table S3**.

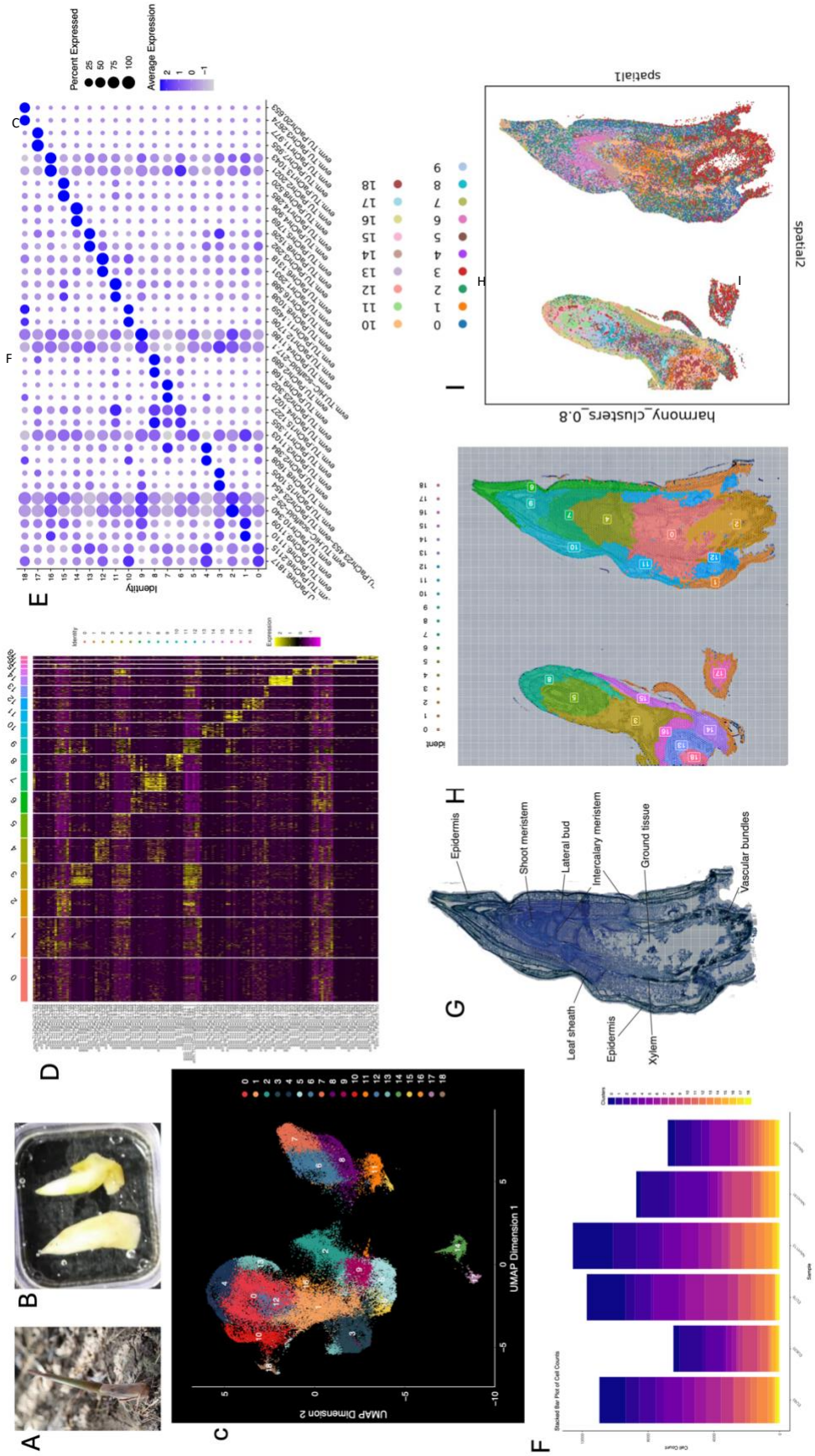

Figure 1 A-B. Tissue from common reed shoots was sampled for single-cell and spatial transcriptomics. C. Uniform Manifold Approximation and Projection (UMAP) plot showing dimensional reduction of six single-cell samples after batch correction using the Harmony method. The projection was derived from 35 principal components (PCs) of highly variable genes. Nineteen distinct clusters are represented by different colors. D. Heatmap displaying the top five gene markers for each cluster. Brighter yellow indicates higher expression levels. E. Bubble plot illustrating the two most reliable markers per cluster. Markers were identified using the FindAllMarkers function in Seurat, applying the ROC test to assess classification power. Markers were ranked by descending ROC value, followed by average log2 fold-change (log2FC) and percentage of presence in the cluster. Darker blue bubbles indicate higher average log2FC values, while bubble size reflects the relative marker expression within each cluster. **F.** Number of cells in each cluster for each sample. Samples EU60, EU620, and EU78 represent non-invasive European populations, while NAint113, NAint119, and NAint61 are from invasive North American populations. **G.** Anatomical structure of the shoot from spatial sample NAint61. The tissue section includes the epidermis, shoot meristem, lateral bud, intercalary meristem, leaf sheath, xylem, vascular bundles, and undifferentiated ground tissue. **H.** Nineteen clusters identified from spatial transcriptome data at a resolution of 37 µm, using level 7 subspots. **I.** Localization of single-cell clusters mapped onto spatial transcriptomic data using the Tangram program. The legend shows the corresponding single-cell clusters. The matching tissue types are discernible from the spatial locations.

Spatial transcriptomics clustering

The anatomical sectioning of the large bud revealed eight distinct cell types on the vertical slice, including epidermis, shoot apical meristem, lateral bud, intercalary meristem, ground tissue, vascular bundles, leaf sheath, and xylem (**Fig. 1G**). The bud on the left provided a different angle of the shoot, with the top part closer to the transection, allowing for the identification of young leaves and vascular tissues. The spatial transcriptomics data, based on level 7 subspots, resulted in 19 clusters (**Fig. 1H**). The single-cell clusters were effectively mapped onto the spatial transcriptomic data, with distinct anatomical structures clearly visible and aligning with their corresponding clusters (**Fig. 1I**).

Cell type determination based on single cell data and projection

To identify the cell types, we performed Gene Ontology (GO) enrichment analysis for each cluster based on known cell markers (**Supplementary Figure S5**). Significant GO terms were identified for clusters 1, 3, 4, 6, 7, 8, 11, 12, 14, 15, 16, and 17. Cluster 0, represents an unknown cell type. The GO enrichment analysis for Cluster 1 revealed biological processes related to responses to oxygen-containing compounds, abiotic stimuli, lipids, and hormones. The spatial projection of these cells suggested that they are mesophylls in aerenchyma, located in the mesophyll and tissues between intercalary meristems. Cluster 2, located below Cluster 1 on the image in leaf sheaths, ground tissue, and epidermis, is likely more developed mesophylls in aerenchyma. Cluster 3, projected to leaf centers, is likely composed of mesophyll cells and associated with various biological processed (**Fig. 2W**). Cluster 4 expressed a classical bundle sheath marker, NADP-ME3, and a vascular marker, AIL5, similar to those found in the *Populus apex*(*18*), suggesting it represents bundle sheath cells. Cluster 5 likely consists of undifferentiated ground tissue, as no specific gene markers were found. Cluster 6, enriched for genes related to very long chain fatty acid and metabolic processes, is projected to the early promeristematic region of the epidermis, indicating it is composed of protoderm cells. Cluster 7, which is actively involved in defense responses to biotic stimuli, stress, metabolic processes, and hormones, is likely to represent leaf guard cells. Cluster 8, enriched in oxoacid metabolic processes and lipid biosynthesis, is likely to be composed of leaf pavement cells. Cluster 9 is projected to mesophylls in the spatial data. Cluster 10, mapped to the xylem and predicted to be xylem by the cell predictor from scplantdb, is confirmed as xylem tissue.

Cluster 11 showed GO enrichment in monocarboxylic acid metabolic processes and was projected to the leaf region. The cell predictor assigned it the highest similarity score to leaf pavement cells. Cluster 12, characterized by response to alcohol, expressed a phloem cell marker, confirming it to be phloem tissue. Cluster 13, predicted to be part of the vascular

system based on the marker gene ATHB-7, is likely a vascular cell type. Cluster 14, projected to the shoot meristematic region, represents meristem cells. Cluster 15, characterized by GO enrichment in the S-adenosylmethionine metabolic process, was projected to the vascular system. Since genes encoding for S-adenosylmethionine are often expressed in lignified tissues(*19*), we classify this cluster as xylem. Cluster 16, enriched in genes related to responses to external biotic stimuli, remains an unknown cell type. Cluster 17, enriched in GO terms related to protein biosynthesis, translation, and organic biosynthesis, and projected to the lateral bud meristem, is classified as lateral bud meristem. Lateral bud meristems were observed in only four individuals: EU60, EU620, NAint61, and NAint191when projecting each sample to the spatial data separately. Finally, Cluster 18, projected to vascular tissues and showing high similarity to companion cells in *Arabidopsis*, is confirmed as companion cells.

The identified cell types were further validated by classical marker genes. In mesophyll cells of Cluster 3, markers for CHLOROPHYLL A-B BINDING PROTEIN CP29.2 (PaLHCB4.2), PHOTOSYSTEM I CHLOROPHYLL A/B-BINDING PROTEIN 1 (PaLHCA2), and PHOTOSYSTEM I REACTION CENTER SUBUNIT III (PaPSAF) were highly expressed, consistent with their known expression in mesophyll cells of *Pisum sativum* (*20*). The epidermal cell-specific marker gene, HOMEODOMAIN GLABROUS 2 (HDG2), was expressed in Cluster 6 (*18*) . Homologs of Arabidopsis PROTODERMAL FACTOR 1 (PDF1) were found in Clusters 8 and 11, which are known as makers of outlayers of epidermis or L1 layer(*18, 20, 21*), confirming these clusters as epidermal cells. Additionally, markers such as HYDROXYSTEROID DEHYDROGENASE 1 (PaHSD1) in Cluster 8, LONG CHAIN ACYL-COA SYNTHETASE 2 (PaLACS2) in Cluster 11, and LONG CHAIN ACYL-COA SYNTHETASE 4 (PaLACS4) in Cluster 7 suggest their involvement in epidermal cell functions (*20*). The ARP gene family, associated with asymmetric leaf development, was identified in Cluster 17, supporting its classification as lateral bud meristem. This gene family, including ASYMMETRIC LEAVES1 (AS1) from *Arabidopsis thaliana*, ROUGH SHEATH2 (RS2) from *Zea mays*, and PHANTASTICA (PHAN) from *Antirrhinum majus* (*22*), codes for MYB-domain transcription factors expressed in lateral organ founder cells (*23*). Additionally, markers indicative of cell proliferation, such as HIGH MOBILITY GROUP B PROTEIN 1 (PaAtHMGB1) in Cluster 14 and UBIQUITIN-CONJUGATING ENZYME E2 19 (PaUBC19) and E2 22 (PaUBC22) in Cluster 17, suggest rapid cell division in these clusters. For vascular tissue, ATHB-5 was highly expressed in Cluster 10 and ATHB-7 in Cluster 13, confirming their involvement in vascular tissue development (*20, 24*). Finally, Cluster 12, characterized by the phloem marker PHLOEM PROTEIN 2-B10 (PP2-B10) (*25*) , is confirmed to be phloem tissue. We projected several strong markers for each single cluster onto the spatial transcriptomic data, as shown in **Fig. 2 A-I**.

RNA velocity and the latent time analyses in all the six single cell samples revealed that cells of cluster 3 are progenitor cells. These cells are labeled with an early latent time, and they have to potential to develop into all the other types of cell and tissues, indicating its multipotency (**Fig. 2 J-K)**. Gene expression from each single-cell cluster was projected onto the spatial transcriptomic data using Seurat, allowing for the identification of the anatomical structure associated with each cluster (**Fig. 2 M-V**).

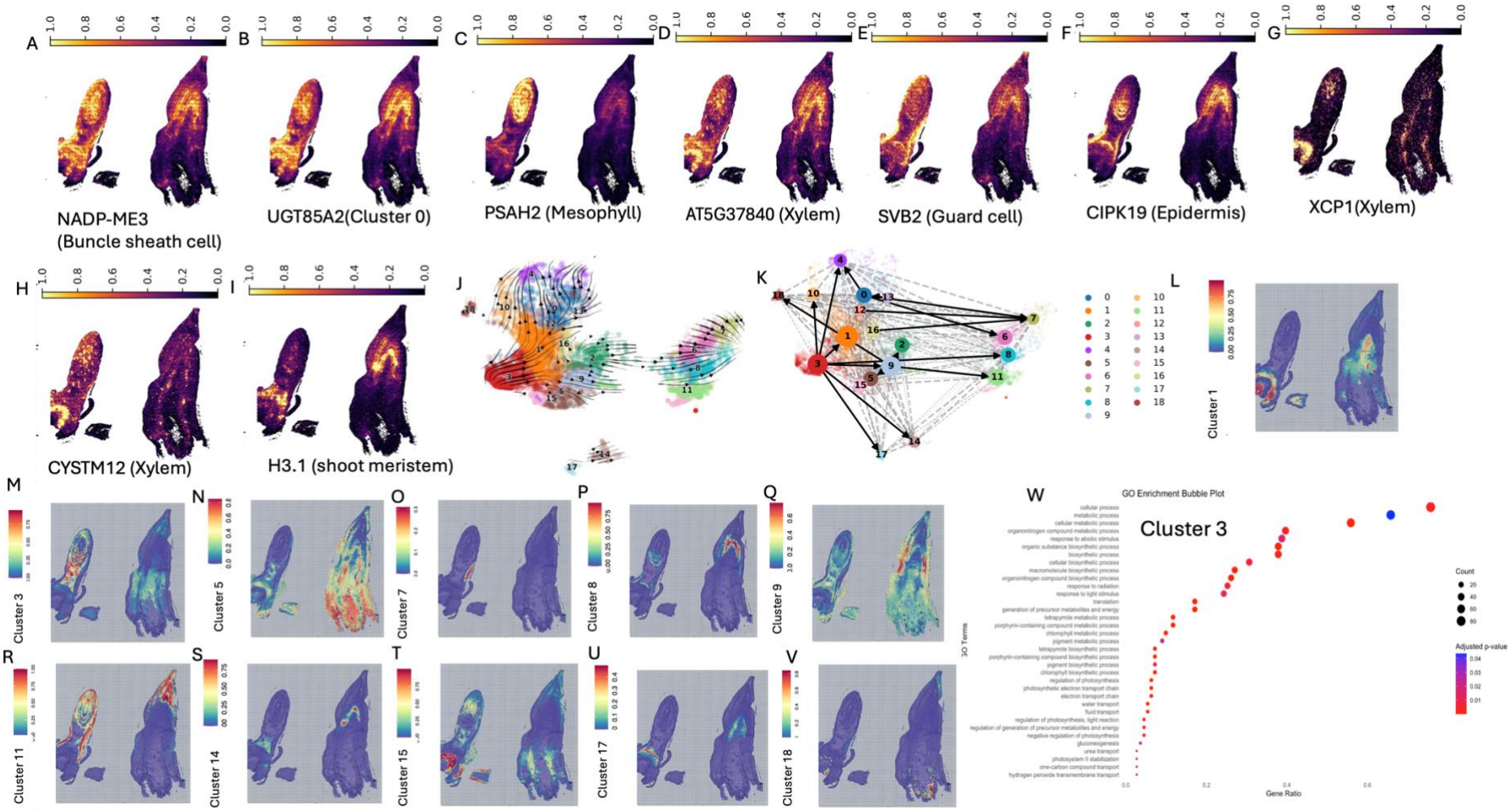
A
NADP-ME3
(Buncle sheath cell)
B
UGT85A2(Cluster 0)
C
PSAH2 (Mesophyll)
D
AT5G37840 (Xylem)
E
SVB2 (Guard cell)
F
CIPK19 (Epidermis)
G
XCP1(Xylem)
H
CYSTM12 (Xylem)
I
H3.1 (shoot meristem)
J
K
L
Cluster 1
M
Cluster 3
N
Cluster 5
O
Cluster 7
P
Cluster 8
Q
Cluster 9
R
Cluster 11
S
Cluster 14
T
Cluster 15
U
Cluster 17
V
Cluster 18
W
GO Enrichment Bubble Plot
Cluster 3
Count
Adjusted p-value
Gene Ratio

Figure 2 A-I. These images display the predicted gene expression on spatial slides for genes identified in the single-cell dataset, representing different cell types. Brighter colors represent higher prediction scores. J. RNA velocity was calculated based on the ratio of mature mRNAs to unspliced RNA transcripts using a dynamical model that incorporates all genes. The arrows indicate the transition states of the cell clusters. H. Partition-based graph abstraction (PAGA) with velocity-directed edges, where edge weights represent the confidence in the presence of connections. L-V. The projection of single cells onto the spatial transcriptomic data using Seurat for clusters 1, 3, 5, 7, 8, 9, 11, 14, 15, 17, and 18. The legend shows the prediction scores of the projections. The remaining clusters were not projected with sufficient confidence. W. Enriched GO terms for cluster 3 in the single-cell dataset. The color represents the adjusted p-value, while the size reflects the number of gene counts.

Differentially expressed genes between EU and invasive lineage

When comparing single-cell sequencing data between the EU and NAint populations, we observed that invasive populations generally exhibited a slightly lower median gene count per cell, suggesting potential differences in cellular transcriptional activity (**Table S2**). Since our focus is on gene regulation in the invasive lineage, we concentrated on the upregulated genes in invasive populations. The highest numbers of upregulated genes were found in clusters 3 and 8, with 2,580 and 2,362 genes, respectively (**Fig. 3E**).

There was no significant GO enrichment in upregulated DEGs in clusters 1, 9, 15, and 16 ($p < 0.01$). For vascular tissues, including clusters 4, 10, 12, 13, and 18, the most common shared GO terms for upregulated DEGs were enriched in responses to heat, hydrogen peroxide, reactive oxygen species (ROS), oxidative stress, hypoxia, organic cyclic compounds, and protein folding. DEGs in cluster 4 were specifically involved in the cellular response to starvation. In cluster 13, DEGs were enriched in processes related to organonitrogen and sulfur compounds, glutathione and monocarboxylic acid metabolism, ribosomal small subunit assembly, cellular modified amino acid metabolism, respiratory electron transport chain, and detoxification.

For the epidermis, including clusters 6, 7, and 8, the common GO terms in upregulated DEGs were enriched in responses to heat and cold, as well as protein folding. In the less differentiated epidermal cells of clusters 6 and 7, overlapping GO terms for upregulated DEGs were enriched in responses to reactive oxygen species. In cluster 8, the GO terms for upregulated DEGs were involved in a range of cellular activities related to metabolism, protein synthesis, energy production, ion response, structural organization, and transport processes, such as macromolecule biosynthesis, cytoplasmic translation, ribosomal small subunit assembly, ribose phosphate metabolism, generation of precursor metabolites and energy, purine ribonucleoside triphosphate metabolism, protein-RNA complex assembly, fatty acid biosynthesis, response to zinc ions, aerobic electron transport chain, cytoskeleton organization, and vesicle-mediated transport. Interestingly, the genes that were downregulated in the invasive lineage were involved in the positive regulation of leaf senescence. In cluster 11, upregulated DEGs were enriched in translation, organonitrogen compound biosynthesis, macromolecule biosynthesis, and protein folding. For the meristematic cells in clusters 14 and 17, shared enriched GO terms mainly involved cytoplasmic translation, responses to cold and hypoxia, organonitrogen compound metabolism, and macromolecule biosynthesis. DEGs specific to cluster 14 included 'de novo' post-translational protein folding, chaperone cofactor-dependent protein refolding, ribosomal small subunit assembly, responses to heat, salt, and oxidative stress, organonitrogen compound metabolism, protein metabolism, carboxylic acid biosynthesis, response to reactive oxygen species, hydrogen peroxide catabolism, translational elongation, and oxoacid metabolism. DEGs specific to cluster 17 were involved in biological processes related to response to acidic chemicals, inorganic substances, water deprivation, osmotic stress, and translation.

In cluster 5, GO enrichment was associated with translation, 'de novo' post-translational protein folding, chaperone cofactor-dependent protein refolding, macromolecule biosynthesis, organonitrogen compound biosynthesis, responses to reactive oxygen species, heterochromatin

organization, and cytoplasmic translation. For DEGs in the chlorenchyma of cluster 3, GO enrichment was linked to core metabolic activities supporting cell survival and growth, including biosynthesis, catabolism, respiration, movement of molecules and macromolecules within cells and across membranes, adaptation to environmental stressors (such as oxygen, salt, osmotic stress, and biotic interactions), and assembly of cellular components and macromolecules to ensure proper function and response. In cluster 2, DEGs in the chlorenchyma were primarily involved in protein folding, protein complex oligomerization, and responses to hydrogen peroxide and reactive oxygen species.

Copy number variation of B chromosome and the expressed genes

After mapping the RAD-seq reads to the reference genome, we observed that invasive individuals consistently exhibited higher read depth than other populations (**Fig. 3A**). This finding was further validated through whole-genome sequencing, which revealed that the depth ratio of B chromosomes to all other chromosomes averaged 1.99 ± 0.335 in the invasive lineages, suggesting an average of four copies of the B chromosome. In contrast, the ratio in the EU population was 0.261 ± 0.0115, indicating a near absence of B chromosomes (**Fig. 3D**).

Three genes on the B chromosomes expressed in every cluster when comparing the EU and invasive lineages (**Fig. 3J-O**). One gene, evm.model.PaChr24.262, homologous to AT4G10890 encoding for DDE family endonuclease, was highly expressed in Cluster 14 and 17, corresponding to the shoot apical meristem and axillary meristems. This suggests that may promote translation or biosynthesis, enhancing regeneration (**Fig. 3F,I**). The AT4G10890 gene exhibited significantly higher expression in the invasive lineage compared to the EU lineage. DDE superfamily genes are involved in coordinating metal ions needed for catalysis, enabling site-specific DNA cleavage and subsequent strand transfer(*26*). Thus, this gene may encode a transposase protein that facilitates the formation of transposable elements in *P. australis*.

Two additional genes located on the B chromosome, evm.model.PaChr24.43 (homologous to IMPA-3) and evm.model.PaChr24.240 (homologous to LOC133886699, encoding a sister-chromatid cohesion protein 3-like protein), also exhibited significantly higher expression in the invasive lineage across all clusters. IMPA-3 is an importin involved in transporting essential proteins into the nucleus. Cohesin is believed to maintain the three-dimensional structure of plant genomes(*27*), particularly during mitosis and meiosis. The coordinated function of these two genes may help prevent chromosomal abnormalities during rapid cell division, promoting enhanced cell proliferation, genome stability, and adaptability to environmental stress.

To investigate the evolution of the B chromosome genes, we looked into the orthologues from phylogeny tree including other Poacae species. The evm.model.PaChr24.43 belongs to OG0000519, and evm.model.PaChr24.240 belongs to OG0003259, and PaChr24.262 belongs to OG0004806. The gene evm.model.PaChr24.43 was clustered with a gene from chromosome 9 (evm.model.PaChr9.309) and chromosome 12(evm.model.PaChr12.264), formed a cluster with a group of importins including AT4G02150 in *Arabidopsis*, Os05t0155601-01 in rice, Zm00001eb346120 in maize, and TKW07008 in *Setaria viridis* (**Fig. 3 BC**). The long branch length suggested mutation occurred frequently on this gene within a short time. Gene evm.model.PaChr24.240 was on the same branch with evm.model.PaChr24.218, all are homologues of evm.model.PaChr10.167 and evm.model.PaChr21.972. Gene evm.model.PaChr24.262 is homologues to evm.model.PaChr17.331. Interestingly, these homologous genes from non-B chromosomes are from the two subgenomes, indicating they are conservative along the evolutionary trajectories.

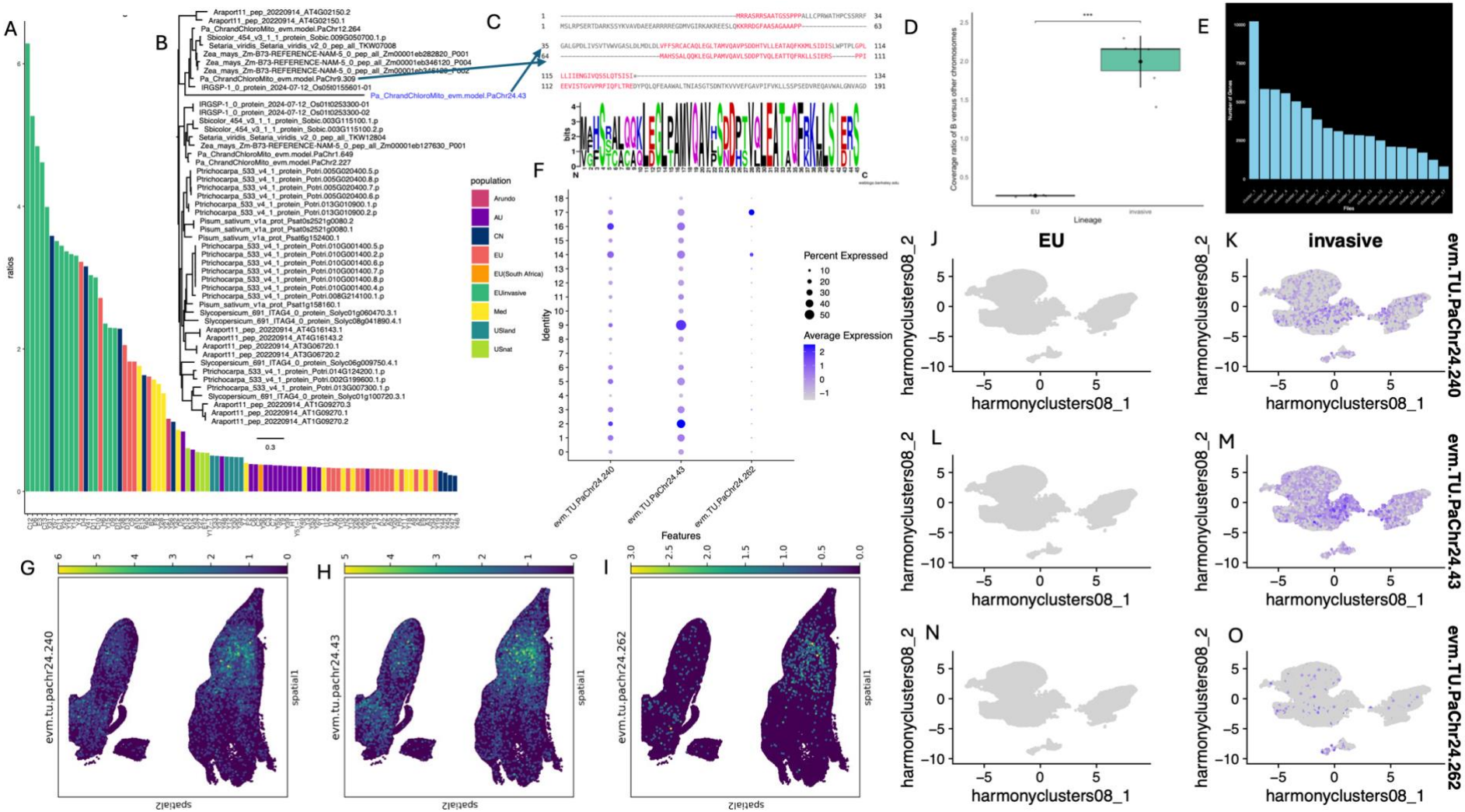
A
ratios
B
population
Arundo
AU
CN
EU
EU(South Africa)
EUinvasive
Med
USland
USnat
C
hits
D
Coverage ratio of B versus other chromosomes
Lineage
EU
invasive
E
Number of Genes
Files
F
Identity
Percent Expressed
Average Expression
Features
evm.TU.PaChr24.240
evm.TU.PaChr24.43
evm.TU.PaChr24.262
G
H
I
evm.tu.pachr24.240
evm.tu.pachr24.43
evm.tu.pachr24.262
spatial1
spatial2
J
EU
K
invasive
L
M
N
O
harmonyclusters08_1
harmonyclusters08_2

Figure 3 A. A bar plot showing the copy number of B chromosomes in different genetic lineages of *P. australis*. Notably, the dark green bars represent invasive individuals in North America, and all these individuals were found to have more copies of B chromosomes than the other chromosomes. B. Phylogeny of orthologues of the B chromosome gene PaChr24.43 (a homologue of IMPA-3) in 10 species, showing a higher mutation rate compared to other genes. C. Alignment of homologous gene pairs evm.model.PaChr24.43 and evm.model.PaChr9.309. The lower panel showed conserved amino acid of evm.model.PaChr24.43 region among the homologous P. australis and Zea mays genes. D. The ratio of reads mapped to the B chromosome versus those mapped to other chromosomes in invasive and non-invasive European populations was compared. The difference was tested using the Mann-Whitney U test. E. The number of differentially expressed genes between invasive and non-invasive population in each cluster. F. The average expression level of the three B chromosome genes in each cluster, estimated using log2FC. Darker blue indicates higher expression, and the size represents the percentage of cells in which the gene is expressed in the cluster. G-I. The spatial projection of the B chromosome gene PaChr24.240, PaChr24.43, PaChr24.262. J-O. The expression level of the three B chromosome genes on non-invasive and invasive individuals.

Growth rates of shoot in the two populations

By comparing the bud diameters of different lineages, it was found that the invasive lineage generally had larger bud diameters and greater variability. Since the bud diameter data for both the invasive and non-invasive lineages did not follow a normal distribution, the Mann-Whitney U test was conducted to determine if there were significant differences between the two lineages. The results revealed a significant difference at the 0.005 level (**Fig. 4A**). Regarding leaf emergence rate, no significant difference was observed in either of the two independent measurements. However, the overall leaf emergence rate was higher in the non-invasive lineage compared to the invasive lineage (**Fig. 4B**).

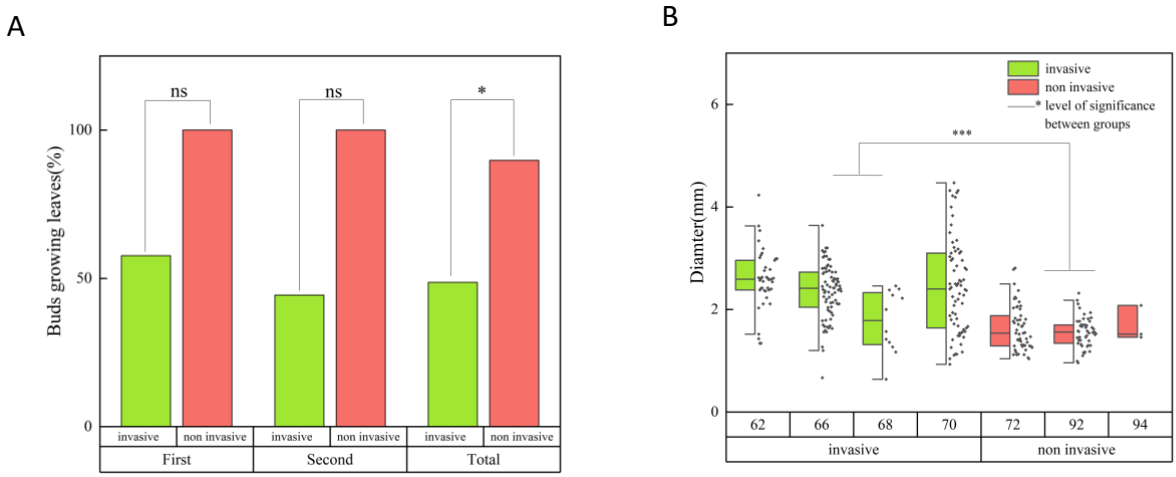

Figure 4 A. The leaf emergence rate of each individual at the conclusion of the experiment, following the fomulla Leaf emergence rate = Number of emerged buds / (Number of emerged buds + Number of dead buds + Number of growing buds). The x-axis represents the grouping information (by lineage and measurement time), with First indicating the first measurement data, Second indicating the second measurement data, and Total representing the combined data from both measurements. The y-axis represents the leaf emergence rate. Each color in the figure represents a different reed lineage. B. The diameter of the bud at the soil surface is measured daily, and the average diameter value is calculated for each bud. The x-axis represents the grouping information (by genotype and lineage), while the y-axis represents the bud diameter values. Each color in the figure corresponds to a different reed lineage. The data points on the right side of each box represent all the measured values for the buds of all individuals within that genotype.

Discussion

Marker detection and cell type determination

This study integrates single-cell and spatial transcriptomic data to investigate the molecular mechanisms underlying shoot development in *P. australis* (common reed). By analyzing gene expression at both the cellular and tissue levels, we aim to uncover key factors that contribute to shoot regeneration and development, particularly in the context of invasive populations. We identified gene markers specific to various tissues, including epidermal cells, the shoot apical meristem, axillary meristem, vascular tissues, and leaves.

Several factors influence the detection of markers across cell clusters in transcriptomic analyses. First, marker detection is more effective when analyzing individual samples rather than clustering multiple samples together. By performing single-sample analysis, we were able to identify a significantly higher number of markers that align with previously published literature, confirming the validity of our approach. Second, we found that clustering cells at lower resolutions in Seurat produces more stable and distinct markers compared to finer clustering strategies. This is evident from the SC3 index stability, which shows that cluster separation is most reliable when the resolution is set between 0.1 and 0.3 (**Supplementary Figure S2**). While higher resolutions can reveal more subtle anatomical features, they also introduce greater variability and less stable marker identification. Third, while most cell types can be accurately predicted by combining classic marker genes with spatial transcriptomic data, some clusters remain ambiguous. This ambiguity is often due to the presence of mixed classical markers representing multiple cell types. It is particularly evident when projections onto spatial slides show diffuse, non-specific patterns of expression. To resolve this ambiguity and accurately identify cell types, it is crucial to rely on multiple genetic markers rather than a single gene. Typically, a set of 50 to 200 genes serves as a reliable pool for distinguishing between different cell types (*28*).

To further validate the functional relevance of our identified markers, we performed BLAST analysis to identify orthologous genes in other species. Interestingly, a significant proportion of these genes showed a greater overlap with maize and sorghum than with *Arabidopsis thaliana*. This is consistent with the evolutionary history of both *P. australis* and maize, both of which have undergone comparable times of whole-genome duplication (WGD) events. These events have resulted in a higher number of distinguishable orthologous genes between these species. In contrast, *Arabidopsis* has experienced at most three common WGD events (*29*), which may explain why orthologues in *P. australis* have diverged in function compared to their counterparts in *Arabidopsis*.

Among the distinct clusters identified, Cluster 1 is predominantly located in tissues between the intercalary meristems. Gene Ontology (GO) enrichment analysis suggests that this cluster functions similarly to mesophyll cells, responding to various environmental stimuli. This suggests that Cluster 1 likely represents mesophyll cells within the aerenchyma, a tissue type specialized for aquatic environments, consistent with the adaptation of *P. australis* to wetland habitats. In contrast, Cluster 4 was characterized by the presence of the marker gene *ATNADP-ME3*, which strongly indicates the presence of bundle sheath cells (*30*).

Chlorenchyma can generate the shoot system

Chlorenchyma is a type of parenchyma tissue in plants that contains abundant chloroplasts and plays a crucial role in photosynthesis by capturing light energy and synthesizing organic compounds. It is typically found in the mesophyll of leaves. Seurat projected Cluster 3 onto the mesophyll in young leaves, while Tangram projected it onto the vascular parenchyma, suggesting potential heterogeneity within this cluster. Gene Ontology (GO) enrichment analysis of Cluster 3 indicates a strong involvement in chlorophyll biosynthesis, organic compound synthesis, and the regulation of photosynthesis, which are functions commonly associated with mesophyll tissue, along with water transport, which is typically related to

vascular tissues. Therefore, this cluster may include both chlorenchyma and vascular parenchyma, reflecting a functional overlap between the two tissue types.

Cluster 3 exhibits the earliest pseudotime in RNA velocity analysis, indicating that it is actively differentiating into various tissue types (**Fig. 2 J-K**). The PAGA graph further suggests that Cluster 3 acts as an initiator for the formation of all cell types, promoting the generation of the shoot apical meristem, epidermis, and subsequent tissues. This observation supports the hypothesis that parenchyma cells may possess totipotency, enabling them to differentiate into an entire plant(*31*). These findings align with a study on tomato callus using spatial transcriptomics, which showed that chlorenchyma cells contribute to shoot system development in tomatoes (*32*). In Poaceae species, it has been observed that younger leaf tissues have a greater regenerative potential than mature tissues when cultured in a cambium-inducing medium, with bundle sheath cells and vascular parenchyma being considered key sources of regeneration(*33*). In our study, the cell markers of Cluster 3 included several chloroplast-related genes, and the cells were projected to the central part of the shoot layer cross-section, reflecting their early differentiation status. This further supports the idea that Cluster 3 cells are crucial in initiating shoot system development and may play a significant role in regeneration processes in *P. australis*.

Programmed cell death in leaf sheath senescence and aerenchyma

In the spatial transcriptomics data, a marker gene, *evm.model.PaChr4.23* (homologous to *AT5G46890*), was found in Cluster 12, which is located at an mature stage of leaf sheath development and is closely associated with programmed cell death (*34*). Several other senescence-related genes were also identified as cell markers in this cluster, including *evm.model.PaChr1.399* (MT2B), *evm.model.PaChr3.27* (AZI3), and *evm.model.PaChr3.28* (AZI3) (34, 35). Another marker, *evm.model.PaChr5.73* (CAT1), encodes an enzyme responsible for catalyzing the reduction of hydrogen peroxide, thus helping to mitigate toxic levels of H_2O_2. Cluster 12 in the spatial data was projected onto the large Cluster 0 in the single-cell data (**Supplementary Figure S6**), where most of the single-cell expression data for Cluster 0 aligns with the epidermal cells in the anatomical structure, particularly in response to reactive oxygen species (ROS). Therefore, Cluster 12 in the spatial transcriptomics data may represent a subset of Cluster 0 in the single-cell data, specifically involved in programming the senescence of epidermal cells starting from the leaf sheath.

This finding is consistent with existing evidence that in monocots, leaf sheath senescence begins from the bottom and progresses upward, as observed in bamboo (*35*), with chloroplast content gradually increasing along the height. In grasses, leaf sheath senescence is characterized by the accumulation of soluble sugars, followed by chemicals such as abscisic acid (ABA), salicylic acid (SA), and H_2O_2 (*35*), which trigger plant responses to stress. The accumulation of ROS is often linked to programmed cell death, supporting the idea that ROS plays a key role in leaf sheath senescence in monocots.

Gene Ontology (GO) enrichment analysis of Cluster 1 in the single-cell data suggests that this cluster may represent mesophyll cells involved in responding to stimuli and stress, likely related to aerenchyma tissue. In the marsh aquatic plant *Typha angustifolia* L. (Typhaceae), the development of leaf aerenchyma is divided into four stages based on morphological characteristics: solid, early cavity formation, late cavity formation, and mature cavity (*36*). Our spatial data identified *APX1* as a cell-type marker for Clusters 4 and 15, which correspond to Cluster 1 in the single-cell data. *Prx52*, a protein similar to peroxidases involved in lignin biosynthesis, was found as a marker for single-cell Clusters 7 and 9, corresponding to spatial Clusters 5 and 15. These findings suggest that Cluster 1 in the single-cell data may represent mesophyll cells of the aerenchyma in the solid stage. This is further supported by the high

expression of *APX* and *Prx* in the solid stage of aerenchyma, with expression levels decreasing in later stages of development (*36*).

B chromosome and its role in plant invasion

The phenotypic measurements suggested that invasive lineages exhibit faster growth in rhizome shoot diameter compared to the European ancestral populations. During the same phenological period in late autumn, the non-invasive lineage continued to grow rapidly and produced more leaves, while the invasive individuals tended to conserve energy for survival. This indicates that the invasive lineage may possess a better mechanism for balancing growth and energy conservation. Structural variations have been shown to play a significant role in shaping invasive genomes. Large haplotypic blocks, including inversions, have been identified as key drivers of rapid adaptation to local environments, potentially serving as a mechanism for the invasion of species such as *Ambrosia artemisiifolia*, as demonstrated by a population genomic study (*37*). Genome scans have revealed a selective sweep in the invasive population, but not in native populations, near the flowering locus *ELF3* in *Ambrosia artemisiifolia.* This region coincides with a haplotypic inversion that preserves adaptive loci, providing a potential advantage for invasive success. Similar mechanisms of structural variation have also been observed in *Helianthus* species (*38*). Hybridization between populations can further enhance adaptability, as structural variation may be transferred to recipient populations, increasing their potential for survival and expansion(*39*). In our study, we identified variations in B chromosome copy number in the invasive population, as well as other genetic lineages, suggesting that these chromosomes may be rapidly accumulating adaptive genes that facilitate survival in novel environments.

The gene *evm.model.PaChr24.43* is homologous to *IMPORTIN-α3/MOS6* (MODIFIER OF SNC1, 6) in *Arabidopsis*, which is a key recruiter of nuclear localization signal-containing cargo proteins into the nucleus (*40*). This gene is known to function as the main receptor of an R (resistance) gene *SNC1* in *Arabidopsis*, playing a critical role in plant innate immunity (*41*, *42*). R genes are often subject to rapid evolution, even within the same species (*43*), interestingly, the importin protein on the B chromosome is also under fast evolution, as evidenced by the phylogenetic tree of this gene in *P. australis*. Nonetheless, its paralogue on Chromosome 9 does not exhibit rapid evolution but instead follows a typical evolutionary rate seen in other species. This suggests that the *IMPA3* gene on the B chromosome is under stronger selection and may be more highly favored in the invasive populations. Additionally, when compared to the homologous gene *evm.model.PaChr9.309*, which contains 531 amino acids, *evm.model.PaChr24.43* is truncated, with only 133 amino acids present, indicating that it represents only a partial version of the gene. This truncated gene may have been translocated to the B chromosome through the activity of transposable elements. Sequence variations in the conserved domain show that *evm.model.PaChr24.43* on the B chromosome has accumulated several amino acid mutations within a relatively short period, further supporting the idea that this gene is undergoing rapid evolution on the B chromosome. Hence, genes on the B chromosomes may be providing an evolutionary advantage in invasive populations by harboring genes that are subject to intensive selection.

Materials and Method

Tissue dissociation and preparation of single-cell suspensions

To generate a comprehensive single-cell and spatial transcriptomic atlas for *P. australis*, we selected six individuals from EU (non-invasive) and NAint (invasive) populations, with three biological replicates in each group. These plants were grown in the common garden of Shandong University under natural conditions. Prior to the single cell protoplast preparation, the plants were kept at room temperature (25 ℃) for a few days, and then the rhizome shoots

were collected. The rhizome shoots were cut into 1-2 mm strips and added to tubes containing 10ml enzyme solution consisting of 2% cellulase R10, 0.8% macerozyme R-10, 0.5% pectinase Y-23, 0.1% BSA, 20mM MES,20mM KCl,10mMCaCL2 and 0.55M mannitol with shaking at 60rpm for 1 hour at 28℃. Cells were then filtered with a 40 um cell strainer. Cell activity was detected by trypan blue staining and cell concentration was measured using a hemocytometer and a light microscope. Lastly, protoplasts were resuspended in 8% (w/v) mannitol solution in preparation for loading onto the chromium controller of the 10xGenomics platform. Approximately, 1.6x104 isolated single cells and enzyme gel-beads were packed into a single oil droplet for single-cell RNA-seq library construction.

Chromium 10x Genomics library and sequencing

Cells were loaded onto the 10X Chromium Single Cell Platform (10X Genomics) at a concentration of 700-1200cells/ul (Single Cell 3′ library and Gel Bead Kit v.3) as described in the manufacturer's protocol. Generation of gel beads in emulsion (GEMs), barcoding, GEM-RT clean-up, complementary DNA amplification and library construction were all performed as per the manufacturer's protocol. Qubit was used for library quantification before pooling. The final library pool was sequenced on the Illumina Nova6000 instrument by Shanghai Personal bio (Shanghai, China) using 150bp paired-end reads.

Spatial transcriptomics sample preparation

We selected NAint61 from the individuals that were used for single cell analysis and performed spatial transcriptomics. A number of fresh shoots are frozen using isopentane and liquid nitrogen bath. Then the frozen tissue samples are embedded in OCT and stored at -80°C. Frozen sectioning is performed using a cryostat, and the tissue sections are collected for RNA extraction and quality control. We selected the samples with a minimum RNA Integrity Number (RIN) value of ≥6 to ensure that the RNA is not degraded during the tissue freezing process. Tissue optimization is performed on the samples before the formal experiment to determine the optimal permeabilization time for spatial gene expression library construction. For the gene expression library preparation, tissue sections are mounted on gene expression slides, followed by methanol fixation, cell segmentation fluorescence staining, cell segmentation fluorescence imaging, H&E staining, and bright-field imaging. Then, tissue permeabilization is performed. The mRNA released from the tissue sections on the gene expression slides is captured by specific probes on the spots and reverse transcribed into cDNA. The cDNA with spatial barcodes is collected from the carrier slides. After double-stranded synthesis, denaturation, PCR amplification, the initial cDNA is completed. Subsequently, enzymatic fragmentation, end repair with A-tailing, magnetic bead fragment selection, adapter ligation, magnetic bead purification, and sample index PCR are performed to construct standard next-generation sequencing libraries. After passing library quality control, the spatial gene expression library is sequenced using a next-generation sequencing platform with a Pair-end 150 sequencing strategy.

Sequencing and data preprocessing of scRNA data

We first demultiplexed the raw reads of scRNA and removed the barcodes of six individuals (EU60, EU78, EU620, NAint61, NAint113, and NAint191). Then the clean reads were mapped to the chromosome level genome assembly(*15*) using cellranger pipeline v8.0 (*44*). The mkref function with default settings in CellRanger was used to build the reference for this species, and count function was used to assess the number of unique molecular identifiers (UMIs) for each gene and barcode. The count matrices were then used for quality control using scanpy(*45*). We only kept genes that are expressed in at least 20 cells, and cells that have at least 500 genes expressed. Cells with mitochondrial genes higher than 10% and chloroplast genes higher than

30% may indicate dead cells thus were removed from the analysis. Doublets were removed using DoubleFinder(*46*). The data of six samples was then integrated into one Seurat object using R package Seurat(*47*), and normalized using SCTransform V2. Since the data were obtained in different batches, we applied harmony method to correct the batch effect after running dimensional redution using PCA. We used 35 PCA axis for further analysis. After trying with different resolution, we selected 0.8 as the final resolution, clustered the cells and run UMAP with parameter min.dist = 0.1, n.neighbors = 50.

Data analysis of spatial transcriptomics

For the spatial transcriptomic data, we used subspot level 7 which includes 127 spots with a resolution of 37μm as a unit to proceed with the downstream analyses. The sequencing data was aligned and quantified using the BMKGENE official software, BSTMatrix. Read 2 was aligned to the reference genome using the STAR software. Cell split was also performed using BSTMatrix, based on a high-resolution image obtained during tissue sectioning. The expression matrix was used as input in Seurat and normalized with SCTransform V2 for downstream analysis, including running PCA, t-SNE(t-Distributed Stochastic Neighbor Embedding), UMAP(Uniform Manifold Approximation and Projection), and marker gene identifications. The clustering was done with a resolution of 0.5.

Cell type annotation and gene markers of single cell clusters

The cell clustering and processing was performed using Seurat R package(*48*). In total, 19 clusters were obtained. For cell type annotation, we first clustered the single cell experiments and projected them to the spatial transcriptomic data to find the matching tissues. Then, we used FindAllMarkers function in Seurat with a threshold of logfc.threshold = 1, min.pct = 0.1, only.pos = TRUE to find all the cell markers for each cluster. We searched for the GO terms of these marker genes to conclude the function of each cluster. Thirdly, we searched for orthologous genes for the cell markers in closely related plant species, including maize, rice, sorghum, and compared cell markers with other literature. To validate the annotation accuracy of the cell types, we projected the cells and clusters of single cell dataset to the spatial transcriptomic slides using both Seurat and Tangram(*49*). Tangram was said to be one of the most accurate software tools for the projection so far, as evidence by a benchmark study(*50*) .

RNA velocity analysis

Following preprocessing steps, the Seurat object of EU620 was converted to h5ad format, and the RNA velocity was evaluated by a velocyte(*51*) run. The loom file containing splicing information was generated with default parameters and RNA velocity was conducted with a python package scvelo(*52*).

Whole genome sequencing

The RADseq dataset were obtained from our previous published work(*8*). We obtained whole genome sequencing data from 10 individuals across Europe and North America. The reads of RADseq and WGS datasets were cleaned and mapped to the reference genome using BWA-mem(*53*). The number of reads mapped to the B chromosome and other chromosomes was calculated using bamcoverage in deeptools(*54*). The differences in ratios between populations were tested using the Student's t test, with $p < 0.05$ considered statistically significant.

Orthologues among species and transcription factor prediction

Eight outgroup species, including *Arabidopsis* (TAIR 11), rice IRGSP1.0 (https://rapdb.dna.affrc.go.jp/download/irgsp1.html)(*55*), *Populus trichocarpa* v4.1(Phytozome)(*18*), sorghum v3.1.1 (Phytozome)(*56*), Pisum sativum(*20, 57*), Solanum

lycopersicum ITAG4.0 (Phytozome)(*58*), maize B73 v5 and *Setaria_viridis* V2 reference genomes at https://plants.ensembl.org/ (*59*) were included to find out the orthologues between species using Orthofinder(*60*). This allows us to obtain genetic markers of each cell type and compare the gene expression between monocots and dicots in the evolutionary trajectories. Transcription factors of the proteins in common reed genome were predicted from PlantRegMap (*61*). The alignment between B chromosome sequences were performed using Cobalt(*62*) in NCBI website, and the visualization of conserved amino acid homologues were conductd using weblogo(*63*).

Plant growth trait measurement

We recorded the daily height of buds above the soil using a ruler and plotted the data by lineage in line charts. The daily growth of buds was calculated by subtracting the height of the previous day from the height of the following day. If the result was negative, it was recorded as 0. These growth values were then plotted as line charts. We measured the diameter of the buds at the soil surface daily using calipers, and the average diameter for each bud was plotted as bar charts by lineage and as box plots by lineage and type. Since multiple data sets did not follow a normal distribution, we performed a Mann-Whitney U test to compare the diameters of invasive and non-invasive types and applied Bonferroni correction for pairwise comparisons of diameters between lineages. Based on the final status of the buds, we calculated the bud sprouting rate, defined as the number of sprouting buds divided by the sum of sprouting buds, dead buds, and buds still growing. A difference analysis by invasive type was conducted using a t-test when variance homogeneity was met; otherwise, Welch's T-test was applied. The results were plotted as bar charts.

Author contributions

C.W., LL. Liu, and WH. Guo conceptualized and supervised the research. LL. Liu and HJ. S prepared the materials for scRNA-seq and spatial transcriptomics. C.W. performed data analysis with contributions from J.O. and MX. Y. HJ. S and LL. Liu analyzed the phenotypic traits. C.W. and J.S. jointly discovered the genomic structural variations. C.W. wrote the manuscript, with contribution from J.O., J.S., LL. Liu, MX. Y, and WH. Guo in the revisions.

Acknowledgement

We thank Dr. Yunfei Cai for insightful discussions on recognizing shoot anatomical structures and Dr. Iivari Kleino for valuable input on single-cell data analysis. We are grateful to the staff at BMK for their comprehensive assistance in obtaining high-quality tissue section slides. We also thank Ms. Wenyi Sheng for her help in caring for the plant materials in the greenhouse. This work was supported by the Natural Science Foundation of Shandong Province (No. ZR2021QC119 to C. Wang and No. ZR2024QC197 to L.L. Liu), the National Natural Science Foundation of China (No. 32470388 to L.L. Liu and U22A20558 to W.H. Guo) and Fundamental Research Funds of Chinese Academy of Forestry (Grant No. CAFYBB2023ZA004 to L.L. Lin). We also acknowledge CSC Finland for their support of computational resources.

1. R. E. V. Sesay, F. Sesay, M. I. Azizi, B. Rahmani, Invasive Species and Biodiversity: Mechanisms, Impacts, and Strategic Management for Ecological Preservation. *Asian Journal of Environment & Ecology* **23**.
2. J. Chen *et al.*, Invasive Spartina and reduced sediments: Shanghai's dangerous silver bullet. *Journal of Plant Ecology* **1**, 79-84 (2008).

3. K. Saltonstall, 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).
4. P. Kaushik, P. K. Pati, M. Khan, P. Khare, Plant functional traits best explain invasive species' performance within a dynamic ecosystem-A review. *Trees, Forests and People* **8**, 100260 (2022).
5. B. Alfaro, D. L. Marshall, Phenotypic variation of life-history traits in native, invasive, and landrace populations of Brassica tournefortii. *Ecology and Evolution* **9**, 13127-13141 (2019).
6. C. E. Lee, Evolutionary genetics of invasive species. *Trends in ecology & evolution* **17**, 386-391 (2002).
7. C. Jaspers *et al.*, Invasion genomics uncover contrasting scenarios of genetic diversity in a widespread marine invader. *Proceedings of the National Academy of Sciences* **118**, e2116211118 (2021).
8. C. Wang *et al.*, Genome-wide analysis tracks the emergence of intraspecific polyploids in Phragmites australis. *npj Biodiversity* **3**, 29 (2024).
9. G. A. Wani *et al.*, Phenotypic variability and genetic diversity of Phragmites australis in Quebec and Kashmir reveal contrasting population structure. *Plants* **9**, 1392 (2020).
10. J. Williams, A. M. Lambert, R. Long, K. Saltonstall, Does hybrid Phragmites australis differ from native and introduced lineages in reproductive, genetic, and morphological traits? *American Journal of Botany* **106**, 29-41 (2019).
11. Y. Zhang *et al.*, Influences of Phragmites australis density and groundwater level on soil water in semiarid wetland, North China: Which is more influential? *Ecohydrology & Hydrobiology* **22**, 85-95 (2022).
12. P. Pyšek *et al.*, Competition among native and invasive Phragmites australis populations: An experimental test of the effects of invasion status, genome size, and ploidy level. *Ecology and Evolution* **10**, 1106-1118 (2020).
13. P. Pyšek *et al.*, Small genome separates native and invasive populations in an ecologically important cosmopolitan grass. *Ecology* **99**, 79-90 (2018).
14. L. A. Meyerson *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).
15. C. Wang *et al.*, Chromosome-level genome assemblies reveal genome evolution of an invasive plant Phragmites australis. *Communications Biology* **7**, 1007 (2024).
16. S. Marschner, K. Kumke, A. Houben, B chromosomes of B. dichromosomatica show a reduced level of euchromatic histone H3 methylation marks. *Chromosome Research* **15**, 215-222 (2007).
17. I. Korsunsky *et al.*, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature methods* **16**, 1289-1296 (2019).
18. D. Conde *et al.*, Single-nuclei transcriptome analysis of the shoot apex vascular system differentiation in Populus. *Development* **149**, dev200632 (2022).
19. I. Sánchez-Aguayo, J. M. Rodríguez-Galán, R. García, J. Torreblanca, J. M. Pardo, Salt stress enhances xylem development and expression of S-adenosyl-L-methionine synthase in lignifying tissues of tomato plants. *Planta* **220**, 278-285 (2004).
20. X. Chen *et al.*, Single-cell transcriptomic analysis of pea shoot development and cell-type-specific responses to boron deficiency. *The Plant Journal* **117**, 302-322 (2024).

21. M. Abe, T. Takahashi, Y. Komeda, Identification of a cis-regulatory element for L1 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein. *The Plant Journal* **26**, 487-494 (2001).
22. M. Tsiantis, R. Schneeberger, J. F. Golz, M. Freeling, J. A. Langdale, The maize rough sheath2 gene and leaf development programs in monocot and dicot plants. *Science* **284**, 154-156 (1999).
23. A. Hasson, T. Blein, P. Laufs, Leaving the meristem behind: the genetic and molecular control of leaf patterning and morphogenesis. *Comptes rendus biologies* **333**, 350-360 (2010).
24. S. R. Moreno, M. O. Lenz, E. M. Meyerowitz, J. C. Locke, H. Jonsson, Single-nucleus transcriptomics resolves differentiation dynamics between shoot stem cells and primary stem. *bioRxiv*, 2024.2008. 2006.606781 (2024).
25. S. Dinant *et al.*, Diversity of the superfamily of phloem lectins (phloem protein 2) in angiosperms. *Plant physiology* **131**, 114-128 (2003).
26. T. Dou *et al.*, Isolation and characterization of a Jerky and JRK/JH8 like gene, tigger transposable element derived 7, TIGD7. *Biochem Genet* **42**, 279-285 (2004).
27. 张雨, 方玉达, Cohesin 结构及功能研究进展. *遗传* **42**, 57-72 (2019).
28. S. Fischer, J. Gillis, How many markers are needed to robustly determine a cell's type? *Iscience* **24**, (2021).
29. C. Simillion, K. Vandepoele, M. C. Van Montagu, M. Zabeau, Y. Van de Peer, The hidden duplication past of Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* **99**, 13627-13632 (2002).
30. B. Zhang, Z. Ma, H. Guo, S. Chen, J. Liu, Single-cell RNA-sequencing provides new insights into the cell-specific expression patterns and transcriptional regulation of photosynthetic genes in bermudagrass leaf blades. *Plant Physiology and Biochemistry*, 108857 (2024).
31. R. Crang, S. Lyons-Sobaski, R. Wise, *Plant anatomy: a concept-based approach to the structure of seed plants*. (Springer, 2018).
32. F. Apelt *et al.*, Shoot and root single cell sequencing reveals tissue-and daytime-specific transcriptome profiles. *Plant physiology* **188**, 861-878 (2022).
33. B. Hu *et al.*, Divergent regeneration-competent cells adopt a common mechanism for callus initiation in angiosperms. *Regeneration* **4**, 132-139 (2017).
34. T. S. Gechev *et al.*, Arabidopsis AAL-toxin-resistant mutant atr1 shows enhanced tolerance to programmed cell death induced by reactive oxygen species. *Biochemical and biophysical research communications* **375**, 639-644 (2008).
35. M. Chen *et al.*, Multi-analysis of sheath senescence provides new insights into bamboo shoot development at the fast growth stage. *Tree physiology* **41**, 491-507 (2021).
36. X. Du *et al.*, Calcium oxalate degradation is involved in aerenchyma formation in Typha angustifolia leaves. *Functional Plant Biology* **45**, 922-934 (2018).
37. P. Battlay *et al.*, Large haploblocks underlie rapid adaptation in the invasive weed Ambrosia artemisiifolia. *Nature Communications* **14**, 1717 (2023).
38. M. Todesco *et al.*, Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* **584**, 602-607 (2020).
39. X. Xia *et al.*, Structural variation and introgression from wild populations in East Asian cattle genomes confer adaptation to local environment. *Genome Biology* **24**, 211 (2023).

40. D. Lüdke *et al.*, Functional requirement of the Arabidopsis importin-α nuclear transport receptor family in autoimmunity mediated by the NLR protein SNC1. *The Plant Journal* **105**, 994-1009 (2021).
41. K. Palma, Y. Zhang, X. Li, An importin α homolog, MOS6, plays an important role in plant innate immunity. *Current Biology* **15**, 1129-1135 (2005).
42. Q. Cai *et al.*, The disease resistance protein SNC1 represses the biogenesis of microRNAs and phased siRNAs. *Nature communications* **9**, 5080 (2018).
43. J. Bergelson, M. Kreitman, E. A. Stahl, D. Tian, Evolutionary dynamics of plant R-genes. *Science* **292**, 2281-2285 (2001).
44. G. X. Zheng *et al.*, Massively parallel digital transcriptional profiling of single cells. *Nature communications* **8**, 14049 (2017).
45. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome biology* **19**, 1-5 (2018).
46. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell systems* **8**, 329-337. e324 (2019).
47. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nature biotechnology* **33**, 495-502 (2015).
48. Y. Hao *et al.*, Dictionary learning for integrative, multimodal, and scalable single-cell analysis (preprint). (2022).
49. A. Regev, T. Biancalani, G. Scalia, L. Buffoni. (Google Patents, 2022).
50. B. Li *et al.*, Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nature methods* **19**, 662-670 (2022).
51. G. La Manno *et al.*, RNA velocity of single cells. *Nature* **560**, 494-498 (2018).
52. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature biotechnology* **38**, 1408-1414 (2020).
53. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25**, 1754-1760 (2009).
54. F. Ramírez *et al.*, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* **44**, W160 (2016).
55. J. Zong *et al.*, A rice single cell transcriptomic atlas defines the developmental trajectories of rice floret and inflorescence meristems. *New Phytologist* **234**, 494-512 (2022).
56. J. Fu *et al.*, Cell-type-specific transcriptomics uncovers spatial regulatory networks in bioenergy sorghum stems. *The Plant Journal* **118**, 1668-1688 (2024).
57. J. Kreplak *et al.*, A reference genome for pea provides insight into legume genome evolution. *Nature genetics* **51**, 1411-1422 (2019).
58. X. Song *et al.*, Spatial transcriptomics reveals light-induced chlorenchyma cells involved in promoting shoot regeneration in tomato callus. *Proceedings of the National Academy of Sciences* **120**, e2310163120 (2023).
59. B. Guillotin *et al.*, A pan-grass transcriptome reveals patterns of cellular divergence in crops. *Nature* **617**, 785-791 (2023).
60. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology* **20**, 1-14 (2019).

61. F. Tian, D.-C. Yang, Y.-Q. Meng, J. Jin, G. Gao, PlantRegMap: charting functional regulatory maps in plants. *Nucleic acids research* **48**, D1104-D1113 (2020).
62. J. S. Papadopoulos, R. Agarwala, COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**, 1073-1079 (2007).
63. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: a sequence logo generator. *Genome research* **14**, 1188-1190 (2004).