Article:
Payet, R.D., Kabir, S.M.T., Morales-Florez, A. et al. (2026) The saltmarsh cordgrass Spartina anglica has novel genomic features that contribute to abiotic stress tolerance. Journal of Experimental Botany. erag363. ISSN: 0022-0957

https://doi.org/10.1093/jxb/erag363

The saltmarsh cordgrass *Spartina anglica* has novel genomic features that contribute to abiotic stress tolerance

Rocky D Payet, Shah Md Tamim Kabir, Alejandro Morales-Florez, Althea M E Rose, Lorelei J Bilham, Matthew E Graham, Charlotte Owen, Lucy S Knowles, Xuan Liu, Muhaiminatul Azizah, Georg Pohnert, Anthony J Davy, Jonathan D Todd, J Benjamin Miller

The saltmarsh cordgrass *Spartina anglica* has novel genomic features that contribute to abiotic stress tolerance

Rocky D. Payet[1,2*ǂ], Shah Md Tamim Kabir[1,2ǂ], Alejandro Morales-Florez[1,2^‡], Althea M. E. Rose[1,2‡], Lorelei J. Bilham[1,2], Matthew E. Graham[1,2], Charlotte Owen[3], Lucy S. Knowles[4], Xuan Liu[5], Muhaiminatul Azizah[6], Georg Pohnert[6], Anthony J. Davy[1], Jonathan D. Todd[1,2,7], J. Benjamin Miller[1,2#]

[1] School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, United Kingdom

[2] Centre for Microbial Interactions, Norwich Research Park, Norwich, NR4 7TJ, United Kingdom

[3] John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, United Kingdom

[4] NERC Environmental Omics Facility, School of Biosciences, University of Sheffield, Sheffield, S10 2TN, United Kingdom

[5] Centre for Genomic Research, University of Liverpool, Liverpool, L69 7ZB, United Kingdom

[6] Bioorganic Analytics, Institute of Inorganic and Analytical Chemistry, Friedrich Schiller University Jena, Lessingstraße 8, 07743 Jena, Germany

[7] MOE Key Laboratory of Evolution and Marine Biodiversity, Frontiers Science Center for Deep Ocean Multispheres and Earth System & College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China

[*] Present address: John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, United Kingdom

[^] Present address: Quadram Institute, Norwich Research Park, Norwich, NR4 7UQ, United Kingdom

[ǂ] These authors contributed equally to this work

[‡] These authors contributed equally to this work

[#] Corresponding author: Dr Ben Miller; b.miller@uea.ac.uk; +44 (0)1603 591627

ORCID IDs:
Dr Rocky D. Payet: 0000-0002-6990-8418
Shah Md Tamim Kabir: 0000-0002-0182-8569
Dr Alejandro Morales-Florez: 0009-0007-0938-0790
Althea M. E. Rose: 0000-0003-2055-824X
Lorelei J. Bilham: 0009-0008-6199-1705
Matthew E. Graham: 0009-0001-0641-1308
Dr Charlotte Owen: 0000-0001-9023-2179
Dr Xuan Liu: 0000-0001-5697-8540
Muhaiminatul Azizah: 0000-0001-8530-414X
Prof Georg Pohnert: 0000-0003-2351-6336
Prof Anthony J. Davy: 0000-0002-7658-7106
Prof Jonathan D. Todd: 0000-0003-0777-9312
Dr J. Benjamin Miller: 0000-0003-0882-033X

Date of resubmission: 29 May 2026

Word count:
Abstract: 199 words
Introduction: 647 words
Materials and Methods: 1899 words
Results: 5077 words
Discussion: 958 words

Figures: 6
Tables: 1
Supporting Figures: 8
Supporting Tables: 26

Running title: *Spartina anglica* osmolytes and genome sequence

KEYWORDS

Spartina anglica, genome, transcriptome, abiotic stress tolerance, osmolytes, saltmarsh, dimethylsulfoniopropionate.

HIGHLIGHT

This work provides genomic and transcriptomic data for the allododecaploid saltmarsh cordgrass *Spartina anglica*, shows that *S. anglica* produces a diverse range of osmolytes, and identifies a retrotransposon-mediated mechanism that causes high-level production of the osmolyte dimethylsulfoniopropionate.

ABSTRACT

The allododecaploid cordgrass *Spartina anglica* thrives in saltmarshes, where it grows vigorously despite facing extreme environmental stresses. *S. anglica* was formed through a genome duplication of an infertile hybrid in ~1870, and is today listed amongst 100 of the World's Worst Invasive Alien Species. Here we use long-read sequencing to provide a genomic sequence of *S. anglica* and undertake a detailed transcriptomic analysis of different *S. anglica* tissues. We demonstrate that *S. anglica* has undergone at least two whole genome duplication events and exhibits segmental allopolyploidy. We also show that *S. anglica* has evolved novel mechanisms to tolerate abiotic stress, including expansion and neofunctionalization of genes involved in sucrose synthesis, detoxification of reactive oxygen species and heavy metals, and calcium signalling associated with salt, ionic and nutrient stress. Furthermore, we demonstrate that *S. anglica* produces a diverse array of osmolytes, including dimethylsulfoniopropionate (DMSP), and reveal that a retrotransposon with a unique 55 bp insertion causes high-level expression of the *DMSP-amine oxidase* (*DOX*) gene and contributes to the exceptionally high levels of DMSP production in *S. anglica*. Together this work highlights the specialised genomic, transcriptomic and biochemical adaptations that *S. anglica* has evolved to tolerate severe abiotic stresses in saltmarshes.

INTRODUCTION

Saltmarshes are significant sites of climate active gas production and are considered carbon sinks (Burden et al., 2019). They are highly saline environments that experience routine inundation and waterlogging by sea water (Hou et al., 2024). Saltmarshes are also frequently contaminated by heavy metals and agricultural run-off (Beeftink et al., 1982), and the low-redox state of anaerobic sediments results in the formation of toxic heavy metal sulfides (Rozema et al., 1991). Consequently, saltmarshes are largely occupied by extremophile plant species that can tolerate this range of abiotic stresses (Veldhuis et al., 2019). Among these species, few are as successful as those of the *Spartina* genus (recently integrated into *Sporobolus*, although not widely accepted (Bortolus et al., 2019)). Many *Spartina* species are highly invasive (Borges et al., 2021), including *Spartina anglica* (also called *Sporobolus anglicus*), and can tolerate salt concentrations up to 600 mM NaCl (Vasquez et al., 2006). Indeed, as of 2023, 68,000 hectares of land in China alone was occupied by *S. alterniflora* (Qi and Chmura, 2023), an extant parent species of *S. anglica*.

S. anglica is an allododecaploid member of the Poaceae family (Ainouche et al., 2004a). This species is thought to have formed in ~1870 as the product of a genome duplication of an infertile hybrid (*Spartina townsendii*), which itself resulted from a cross between the hexaploid species *Spartina alterniflora* and *Spartina maritima* (Ainouche et al., 2004a). Considered a xenonative to the United Kingdom, *S. anglica* has since been introduced to Asia, Australia, New Zealand and North America (Bortolus et al., 2019), where it is also considered a highly invasive species and an “ecosystem engineer”. As a result, *S. anglica* is now listed by the Invasive Species Specialist Group as one of the “100 of the World’s Worst Invasive Alien Species” (GISD), highlighting its global significance.

The halotolerance of *Spartina* species can be attributed to specialist adaptations such as the formation of metal-recreting salt glands (Burke et al., 2000), high-level production of a diverse suite of osmolytes such as glycine betaine (GB) and dimethylsulfoniopropionate (DMSP) (Payet et al., 2024), and the putative expansion of gene families encoding transporters involved in salt stress (Yang et al., 2023). Additionally, *S. anglica* can grow in otherwise anaerobic sediments whose rhizosphere redox potential is elevated by the

diffusion of oxygen from aerenchyma in its roots that provide a conduit from the atmosphere (Granse et al., 2022). These specialist adaptations are thought to have contributed to the success of *S. anglica* as an invasive species.

Despite the ecological significance of *S. anglica* and its remarkable abiotic stress tolerance, no genomic resources and virtually no transcriptomic resources exist for this species. Furthermore, to date, few genomes of saltmarsh-inhabiting plant species have been sequenced, and so the effect on a genome of evolving in such an environment remains largely unexplored. Here, using long-read sequencing, we present the first genome sequence for *S. anglica* and evidence for segmental allopolyploidy in this species. We also provide transcriptomic sequencing data for five different *S. anglica* tissue types, including leaf, stem, inflorescence, root, and rhizome. We show that *S. anglica* has evolved novel mechanisms to help it adapt to the severe environmental stresses of saltmarshes, including expansion and neofunctionalization of genes involved in sucrose synthesis, detoxification of reactive oxygen species and heavy metals, and calcium signalling associated with salt, ionic and nutrient stress. We also demonstrate that *S. anglica* produces a diverse array of osmolytes, including GB, DMSP, proline, proline betaine (stachydrine), ectoine and dimethylsulfoxonium propionate (DMSOP). Furthermore, we demonstrate that DMSP production occurs at extraordinarily high levels in *S. anglica* in part because of a retrotransposon with a unique 55 bp insertion that causes high-level expression of the *DMSP-amine oxidase* (*DOX*) gene in the DMSP synthesis pathway. Together this work highlights the various specialised adaptations that *S. anglica* has evolved to tolerate the harsh abiotic stresses of the saltmarsh environment in which it grows.

MATERIALS AND METHODS

Plant material and genome sequencing

S. anglica material was collected from a clonal clump at Stiffkey saltmarsh, Norfolk, UK, 2021. Material was washed with deionised water to remove surface contamination and then snap frozen in liquid nitrogen. For high molecular weight DNA extraction, plant material was ground under liquid nitrogen to a fine powder. 15 g of this powder was

provided to the NERC Environmental Omics Facility (NEOF), UK, where high molecular weight DNA extractions were performed using the Nucelobond HMW kit (Macherey-Nagel, 744160.1). Lysis buffer and proteinase K volumes were doubled, compared with the kit's protocol. The sample was incubated for one hour with gentle agitation every 15 minutes. RNase A and binding buffer volumes were also doubled and the rest of the process completed per the kit instructions. PacBio libraries were prepared using the SMRTbell prep kit 2.0 (PacBio, 100-938-900). Libraries were sequenced using a PacBio Sequel2e platform (PacBio, USA) over 4 SMRT cells, resulting in a total of 6,257,615 HiFi PacBio reads.

RNA extraction and sequencing

S. anglica plant material was sampled as above and washed with deionised water to remove surface contamination. Plant material was dissected into leaf, stem, inflorescence, root, and rhizome tissue, in triplicate. Following this, material was ground to a fine powder in liquid nitrogen. RNA was extracted from this powder using Tri-Reagent solution (ThermoFisher Scientific, AM9738) as per the manufacturer's instructions. RNA was cleaned using Zymo RNA Clean Concentrator^{TM}-25 (Zymo-Research). Extracted RNA samples were sent to Novogene Co. where RNA-seq libraries were prepared and sequenced with paired-ends to an average depth of 20 million reads per sample. For quantitative RNA-seq analysis, adapter sequences were first trimmed using Trim Galore (Martin, 2011). Following this, Salmon (Patro et al., 2017) was used to quantitatively map reads to the coding sequences of our *S. anglica* genome annotation. Differential gene expression was then calculated using DESeq2 (Love et al., 2014), with an adjusted p-value of $p \leq 0.05$. To visualise expression differences, DESeq2 normalised read numbers were used.

Genome assembly

The genome assembly was generated by NEOF, UK. Briefly, HiFi PacBio data was assembled using HiFiASM and HiCanu assuming a haplotig number of 6, since *S. anglica* is a hexaploid species that underwent genome duplication (Ainouche et al., 2004b) and has an estimated haploid genome size of 438 Mb (Baumel et al., 2002). This resulted in

an assembly of length 2,978,325,964 bp, containing 497 contigs with an N50 of 48,744,091 bp. Contamination was assessed using Blobtools2 (Challis et al., 2020) with default settings; raw read mapping confirmed uniform coverage at a mean depth of 30.43x. Blobtools2 taxonomy assignment returned predominantly no-hit sequences; BLASTn querying of 2,000 subsampled reads against the NCBI nt database resolved these as phylogenetically related Poaceae (*Oryza sativa*, *Zea mays*, *Panicum* spp., *Sporobolus maritimus*, *Spartina alterniflora*), and were consistent with the GC content (~45%) of the target organism. No non-plant contaminants were detected at levels warranting removal.

Gene annotation

The assembled *S. anglica* genome sequences were annotated using MAKER v3.01.03 (Cantarel et al., 2008) as previously described (Card et al., 2019). The initial step involved constructing a *de novo* repeat library using RepeatModeler v2.0.3 (Flynn et al., 2020) and RepeatMasker v4.1.1 (Nansheng, 2004). Subsequently, Trinity v2.11.0 (Grabherr et al., 2011) was employed for the *de novo* assembly of RNA-seq data obtained from the various plant tissues (leaf, stem, inflorescence, root, and rhizome), serving as EST evidence. Protein sequences from *Brachypodium distachyon*, *Panicum hallii*, *Arabidopsis thaliana* and *Oryza sativa* were retrieved from the UniProt database for protein homology evidence. Three assemblers were trained using the initial round of maker outputs: Augustus (Stanke et al., 2006), SNAP (Korf, 2004) and GeneMark HMM (Lukashin and Borodovsky, 1998). The gff3 files derived from the first round MAKER output, along with the trained data, were utilised as inputs for the second round. Custom Perl scripts were employed to extract CDS and protein sequences from the second round gff3 file. The genome completeness was evaluated using BUSCO v5.4.7 (Simão et al., 2015) with embryophyte (odb10) datasets. Filtering was then performed to retain only those sequences which had a minimum of one transcript per million (TPM) in RNA-seq data generated from all replicates of at least one tissue-type. The functional annotation of protein coding genes was performed by BLASTp with UniProt/Swiss-Prot proteins (The UniProt Consortium, 2025), NCBI-NR, InterProScan v5.63-95.0 (Jones et al., 2014), and

Gene Ontology (GO) as defined by EggNOG (Huerta-Cepas et al., 2019). All of these annotations were merged into a combined gff3 file.

Orthogroups and gene family duplication

Proteins from the prediction pipeline were grouped into orthogroups using Orthofinder (Emms and Kelly, 2019). The proteomes of *Eragrostis curvula*, *Musa acuminata*, *Zea mays*, *Sporobolus stapfianus* and *Sporobolus pyramidalis* were used for grouping. Gene family duplication was then computed using CAFE (Computation Analysis of gene Family Evolution (De Bie et al., 2006) using default parameters, with the phylogenetic species tree and orthogroup count data provided as inputs. Duplication events were then filtered so that only those with an associated p-value of 0.05 or lower were considered. This list of duplications was then manually examined, and those orthogroups that were implicated in abiotic stress were taken forwards for more detailed analyses.

Whole genome duplication analysis

S. anglica protein sequences and CDS nucleotide sequences were provided to both WGDdetector (Yang et al., 2019) and MCScanX (Wang et al., 2012). Both of these tools were run using default parameters and Ds and Ks values were retrieved, respectively. These values were then plotted against density and the resulting plots were used to infer genome duplication events.

Constructing phylogenetic trees

Orthofinder (Emms and Kelly, 2019) was used to establish the phylogeny of *Spartina anglica*. This makes use of the Species Tree Root Inference from Duplication Events (STRIDE) algorithm, which takes all identified orthogroups as an input, constructs gene trees, and then infers the root using duplication events within these.

For phylogenetic trees of specific genes, sequence alignments were performed using MAFFT (Katoh and Standley, 2013) with a maximum of 1000 iterations and the FFT-NS-i model. The resulting alignments were then used as input IQTree3 (Wong et al., 2025), using 1000 ultrafast bootstraps optimised by nearest neighbour interchange, and substitution model selection by ModelFinder (Kalyaanamoorthy et al., 2017).

GO enrichment analyses

Gene ontology enrichment analysis was performed at a tissue level using TopGO (Alexa et al., 2006). The corresponding "Gene Universe" was constructed using the annotated genes.

Measurement of osmolytes by ultra high-performance liquid chromatography-high resolution mass spectrometry (UHPLC-HRMS)

Intact leaves from the three highest and three lowest DMSP-accumulating *S. anglica* plants from Stiffkey saltmarsh were sampled and snap frozen in liquid nitrogen. Approximately 5 mg of dry-weight leaf material was extracted with 500 µL of LC-MS grade methanol and incubated for 30 minutes. After centrifugation (for 15 min at 16,100 *g*), three aliquots of 5 µL of each supernatant were diluted with 95 µL of a mixture of acetonitrile and water (9:1 v/v) containing all internal standards (D_6-DMSP, D_6-DMSA, D_3-gonyol, and D_3-ectoine) with a final concentration of 5 nM. This solution was subjected to UHPLC-HRMS analysis.

The analysis of osmolytes via UHPLC-HRMS was based on a previously reported protocol (Azizah and Pohnert, 2022, Azizah and Pohnert, 2023). Separation was performed on a Sequent ZIC-HILIC column (2.1 × 150 mm, 5 µm) equipped with a SeQuent ZIC HILIC guard column (2.1 × 20 mm, 5 µm) (Merck, Darmstadt, Germany). UHPLC mobile phase consisted of high-purity water (Th Geyer GmbH, Renningen, Germany) with 2% LC-MS grade acetonitrile (Th Geyer GmbH, Renningen, Germany) and 0.1% LC-MS grade formic acid (Thermo Scientific, Dreieich, Germany) (solvent A), and 90% acetonitrile with 10% aqueous LC-MS grade ammonium acetate (1 mM) (LGC Promochem, Wesel, Germany) (solvent B). A gradient elution was performed using an isocratic elution of 100% solvent B for 1 min, followed by a linear gradient from 100% solvent B to 20% solvent B over 5.5 min, then a linear gradient from 20% solvent B to 100% solvent B over 0.6 min, and finally an isocratic re-equilibration at 100% solvent B for 2.9 min. The total run time was 10 min. The column was kept at 25 °C. The flow rate

was set at 0.6 mL min^{-1}. The injection volume was 5 µL. Full scan mode was set from 75 to 200 *m/z* at a resolution of 70,000.

Identification of osmolytes was carried out by comparing retention time and MS spectra of osmolytes in the samples with commercial and synthetic standards. Commercially available standards, stachydrine (Pytolab GmbH, Dutendorfer, Germany), choline (Sigma-Aldrich, St. Louis, MO, USA), sarcosine (ABCR GmbH, Karlsruhe, Germany), ectoine (Sigma-Aldrich, St. Louis, MO, USA), and proline (Sigma-Aldrich, St. Louis, MO, USA) were used. DMSP and DMSOP were obtained by synthesis as described in previous studies (Gebser and Pohnert, 2013, Thume *et al.*, 2018).

For quantification, a calibration curve of the standards was recorded, followed by comparisons of the peak area of the analyte with that of the internal standard. As internal standards for quantification of DMSP, DMSOP, and ectoine, we used D_6-DMSP, D_6-dimethylsulfonioacetate (DMSA), and D_3-ectoine, respectively. For the rest of the osmolytes, we used D_3-gonyol as the internal standard. D_6-DMSP, D_6-DMSA, D_3-gonyol, and D_3-ectoine were acquired by synthesis in our laboratory based on published procedures (Gebser and Pohnert, 2013, Thume *et al.*, 2018, Fenizia *et al.*, 2020). For DMSOP, the calibration curve (n = 3) for the area of the molecular ion was $y = 6.75 \times 10^{-2} x$ with r = 0.9944, limit of detection (LOD) = 0.6 nM, limit of quantification (LOQ) = 2.3 nM; for DMSP, the calibration curve (n = 3) for the area of the molecular ion was $y = 2.38 \times 10^{-1} x$ with r = 0.9944, limit of detection (LOD) = 0.6 nM, limit of quantification (LOQ) = 20.1 nM; for ectoine, $y = 2.49 \times 10^{-1} x$ with r = 0.9962, limit of detection (LOD) = 2.5 nM, limit of quantification (LOQ) = 8.8 nM; for stachydrine $y = 2.03 \times 10^{-1} x$ with r = 0.9860, limit of detection (LOD) = 19.1 nM, limit of detection (LOD) = 66.1 nM; for choline, $y = 4.61 \times 10^{-3} x$ with r = 0.9968, limit of detection (LOD) = 55.0 nM, limit of quantification (LOQ) = 193.8 nM; for glycine betaine, $y = 3.85 \times 10^{-1} x$ with r = 0.9890, limit of detection (LOD) = 2.2 nM, limit of quantification (LOQ) = 7.4 nM; for proline $y = 4.69 \times 10^{-3} x$ with r = 0.9896, limit of detection (LOD) = 57.1 nM, limit of quantification (LOQ) = 200.6 nM; for sarcosine $y = 2.31 \times 10^{-3} x$ with r = 0.9887, limit of detection (LOD) = 66.3 nM, limit of quantification (LOQ) = 231.9 nM. All calibration curves passed Mandel's test.

Testing LTR^{DOX} function in *Nicotiana benthamiana*

Plasmids containing the 35S minimal promoter, with and without LTR^{DOX}, and driving expression of the firefly luciferase (*LUC*) reporter gene (LTR^{DOX}-p35Sminimal-LUC-t35S and p35Sminimal-LUC-t35S) were constructed, and then assembled into multi-gene constructs containing a constitutively expressed β-glucuronidase (*GUS*) reporter gene cassette (pAtUBI10-GUS-tNOS), using Golden Gate cloning, as described previously (Tansley et al., 2023). The sequence-verified plasmids were transformed into *Agrobacterium tumefaciens* strain GV3101 and transformation of *N. benthamiana* leaves was performed in triplicate by infiltration as previously described (Tansley et al., 2023). Three days after infiltration, the transgenic leaf material was harvested and luciferase and GUS activities were quantified, as previously described (Feike et al., 2019), using a Hidex Sense microplate reader.

RESULTS

Sequencing, assembly and annotation of *S. anglica* genome

To generate a high-quality genome assembly of *S. anglica*, a total of 6.26 million HiFi reads consisting of ~91.6 (~30.8x) Gb were generated using a PacBio Sequel2e platform (Supplementary Table S1). The draft genome assembly length was 2.97 Gb in 497 contigs (N50 = 48.74 Mb) (Table 1). The expected haploid chromosome number of *Spartina anglica* is 62 (Renny-Byfield et al., 2010), and, consistent with this, the largest 62 contigs of the 497 contigs generated contained 2.89 Gb and represented 96.9% of the total assembled genome (Supplementary Table S2). BUSCO was run on this assembly and reported a 97.7% completeness score and 85.5% duplication.

Repeat annotation revealed that 71.3% of the *S. anglica* genome is composed of repeating elements (Table 1; Supplementary Table S3). Most of these repeating elements were made up of retrotransposons (36.4%), which primarily consisted of long terminal repeat (LTR) retrotransposons of the Ty3/gypsy and Ty1/copia types (23.7% and 10.2%, respectively). The remaining repeating elements comprise a combination of long interspersed nuclear elements (LINEs; 2.3%), DNA transposons (3.9%), rolling-circle

transposons (0.2%) and repeat elements which could not be classified (31.0%) (Supplementary Table S3). These values for *S. anglica* are consistent with those from the recently published genomes of *S. alterniflora* and *S. maritima* (Salmon et al., 2025).

Annotation of the *S. anglica* draft genome was performed using MAKER (Cantarel et al., 2008), combining *de novo* homology-based evidence and transcriptome data from five different tissue types. MAKER was run twice and filtered based upon evidence of expression, resulting in a total of 91,935 genes which correspond to 369.6 Mb (12.4%) of the genome. The average gene length (introns and exons) was 4,021 bp. The number of noncoding tRNA and rRNA were 1,517 and 6,959, respectively (Supplementary Table S4). BUSCO analysis resulted in 92.8% completeness, using embryophyta_odb10 as a database (Supplementary Table S5). Of these gene predictions, 82,083 were functionally annotated using EggNOG and the NCBI non-redundant database, and 65,259 and 69,849 were assigned using UniProt and InterProScan, respectively (Supplementary Table S6). All annotated genes were restricted to only 101 contigs of the assembled 497, with Sa007 demonstrating the highest number of genes at 3,649 (Figure 1A; Supplementary Table S7).

Collinearity analysis revealed that genes in 68 contigs exhibit collinearity. In total, 69,319 genes are involved in segmental duplication, forming 108,416 collinearity pairs, indicating that 75% of genes evolved from whole-genome duplication. Additionally, 1,086 genes had undergone tandem duplication, remaining in close proximity, with 1,323 tandem duplication events recorded. A total of 13,178 dispersed genes evolved through transposition events, including replicative, non-replicative and conservative transpositions. The dataset also included 7,592 single-copy (singleton) genes (Supplementary Table S8).

Spartina anglica has undergone at least two whole genome duplication events and exhibits segmental allopolyploidy

A phylogeny of *S. anglica* was constructed using a combination of genomes representing both the PACMAD and BOP clades of *Poaceae*, and the sister *Musaceae* family (Figure

1B). The placement of *S. anglica* is consistent with it belonging to the PACMAD clade and to the Chloridoideae subfamily. *S. anglica* was recently subsumed into the genus *Sporobolus*, a decision which has divided scientific opinion (Bortolus et al., 2019). Recently, two established *Sporobolus* genomes were sequenced (Chavez Montes et al., 2021) and were thus included in this analysis. These two species, *Sporobolus stapfianus* and *Sporobolus pyramidalis*, were more closely related to one another than to *S. anglica*, consistent with *Spartina* being a coherent clade within the genus *Sporobolus* in the new taxonomy.

The genome assembly was next examined for evidence of genome duplication events using two separate tools, MCScanX (Wang et al., 2012) and WGDdetector (Yang et al., 2019), which yielded different but compatible results. Analysis of the Ks value distribution of the assembly using MCScanX showed a clear peak at approximately Ks = 1.25 (Supplementary Figure S1A), consistent with an ancestral whole genome duplication (WGD) event. There was evidence of this peak in Ks values generated from WGDdetector, although the peak was less pronounced (Supplementary Figure S1B). Interestingly, the data generated with WGDdetector suggested a recent WGD (Ks = 0.35; Supplementary Figure S1B), which was not apparent in the MCScanX results. Taken together, these data suggest that *S. anglica* has undergone at least two WGD events, one recent and one ancestral, and that these two events are distinct from the reported WGD event undergone in *S. townsendii*. Two genomes for the parental species *S. alterniflora* were recently published (Chen et al., 2024; Hao et al., 2024). The Ks values for *S. anglica* obtained with MCScanX (Supplementary Figure S1A) are consistent with published Ks values derived from the *S. alterniflora* genome, supporting a recent WGD event. However, this was not true of the values we obtained using WGDdetector, which suggested a more ancient WGD event. It is therefore possible that the more ancient WGD seen clearly in Supplementary Figure S1B arises from the *S. maritima* genome, and, indeed, a recent study reporting the *S. maritima* and *S. alterniflora* genomes supports this notion (Salmon et al., 2025).

The Chloridoideae subfamily of grasses contains the genus *Sporobolus* and is predominantly polyploid. Microsynteny analysis revealed that *Spartina anglica* exhibited a 6:1 pattern with the diploid grass species *Oropetium thomaeum*. In contrast, two *Sporobolus* species displayed 2:1 and 3:1 patterns with *O. thomaeum* (Figure 1C). Macrosynteny analysis further highlighted the polyploid nature of *Sporobolus stapfianus* and *Sporobolus pyramidalis*, showing 2:3 patterns that clearly indicated their tetraploid and hexaploid states, respectively (Supplementary Figure S2), as previously documented (Chávez Montes et al., 2022). The syntenic block analysis demonstrated that 79% of *O. thomaeum* blocks shared a 1:6 pattern with *Spartina anglica*. Approximately half of *Sporobolus stapfianus* blocks shared a 1:6 pattern, while 34% followed a 2:6 pattern with *Spartina anglica*. Similarly, *Sporobolus pyramidalis* showed diverse syntenic relationships, with 25% of its blocks following a 3:6 pattern, nearly 50% aligning with a 2:6 pattern, and 17% showing a 1:6 pattern with *Spartina anglica* blocks (Supplementary Figure S2). These macrosynteny results suggest a more intricate polyploid structure than the straightforward diploid, tetraploid, hexaploid, or dodecaploid classifications. The findings imply that these species, including *S. anglica*, may exhibit segmental allopolyploidy, with varying degrees of homology among chromosomes derived from different subgenomes.

Gene expression in *S. anglica* is highly influenced by abiotic stresses associated with the saltmarsh environment

As part of the annotation pipeline, we performed RNA-seq on leaf, stem, inflorescence, root, and rhizome tissues from environmentally derived samples of *S. anglica*. The RNA-seq data was then quantitatively mapped to gene annotations to investigate the gene expression profiles for each of these tissues. A principal component analysis demonstrated that the transcriptomes of each of the five tissues were distinct (Supplementary Figure S3), although the root and rhizome tissues clustered together closely. The top 100 most highly abundant transcripts of each tissue were subset and a gene ontology (GO) enrichment analysis for biological process terms was performed on these genes. The top 20 most highly enriched terms were then examined (Supplementary Table S9). Unsurprisingly, the most abundant transcripts in leaf tissue pertained to

photosynthesis. This was also true of inflorescence tissue, despite these two tissues clustering distinctly from one another (Supplementary Figure S3). There was also a substantial enrichment in inflorescence tissue for the biosynthesis and metabolism of polyamines, such as spermine, which are implicated in developmental processes and environmental stress responses (Blázquez, 2024). Among these top 100 genes in each tissue, many pertained to abiotic stress, with GO process terms such as: response to oxidative stress, response to water deprivation, and response to temperature (Supplementary Table S9; Supplementary Table S10). Indeed, when the GO process terms of the top 100 transcripts of each tissue were compared, the eight enriched process terms common to all five tissues were: response to abiotic stimulus, response to inorganic substance, response to light stimulus, response to radiation, response to water deprivation, response to water, response to chemical and response to toxic substance (Supplementary Figure S4; Supplementary Table S10). Taken together, these data suggest that the transcriptome of *S. anglica* is heavily influenced by abiotic stresses in the environment, particularly those stresses that influence the water status of the plant.

Expansion of gene families in *S. anglica* likely contributes to tolerance of physiological stress in the saltmarsh

Over the course of evolutionary time, gene families are subject to expansions and contractions, which, when acted upon by selection, can play a role in adaptation (Demuth and Hahn, 2009). Saltmarshes are high stress environments, as reflected in the transcriptomic profiles of the various *S. anglica* tissues (Supplementary Figure S4; Supplementary Table S9; Supplementary Table S10). We therefore hypothesised that the abiotic stresses encountered by *S. anglica* in its native environment may constitute selection pressures sufficient to promote expansion of gene families associated with abiotic stress tolerance. To analyse these events, we ran CAFE (De Bie et al., 2006) using all orthogroups identified between *Spartina anglica, Eragrostis curvula*, *Musa accuminata*, *Zea mays*, *Sporobolus stapfianus*, and *Sporobolus pyramidalis*. This revealed 159 gene families showing expansion or contraction in the *S. anglica* genome (Supplementary Table S11; Supplementary Table S12), with an associated p-value of 0.05 or lower. Of these, 149 families had gained members (Supplementary Table S11) and 10

had lost them (Supplementary Table S12), although this latter category should be treated with caution as the BUSCO score was 93%, indicating some genes may be missing from our assembly. Furthermore, there was a positive correlation between gene expression level and number of gene family members (Supplementary Figure S5, R^2 = 0.2878), suggesting that gene family expansion facilitated greater levels of transcription, which could provide a fitness advantage. Amongst these expanded gene families, some of the greatest increases in size were observed in abiotic stress responsive families (Supplementary Table S11). For example, the sucrose synthase (SUS) family, involved in hypoxia, cold, flooding, and osmotic stress (Wang et al., 2014; Xu et al., 2019; Gurrieri et al., 2020) had doubled in size (10 members in total), and the heat shock protein 70 family (HSP70), involved in protein folding, temperature stimulus, and response to metals (Usman et al., 2017; Jiang et al., 2020) had also increased (18 members in total; Supplementary Table S11). Furthermore, of the top 100 most highly expressed transcripts in each tissue (Supplementary Table S10), *SUS* and *HSP70* family members were common to all five tissue types. In addition to the *SUS* and *HSP70* families, a number of other gene families involved in stress responses also appear to have undergone expansion, including salt sensitivity *NST1-like* genes (Mitsuda et al., 2005), *HSP20* genes (Ji et al., 2019; Guo et al., 2020; Lian et al., 2022), chloride channels (Liu et al., 2023), and universal stress protein family members (Chi et al., 2019) (Supplementary Table S11). Overall, expansion of these abiotic stress-associated gene families likely represents important innovations that together contributed to the adaptation of *S. anglica* to physiological stress tolerance in saltmarshes and have allowed this species to colonise such an inhospitable environment.

The sucrose synthase (SUS) family is significantly expanded in *S. anglica* and shows evidence of neofunctionalization

In plants, the sucrose synthase (SUS) protein family falls into three main clades (Stein and Granot, 2019), with members of clade I being stress-responsive and proposed to be the most functionally important (Xu et al., 2019; Gurrieri et al., 2020). One orthogroup of SUS family genes had undergone an expansion in *Spartina anglica* (p = 0.011), having gained 5 members relative to closely related species (Supplementary Table S11). To

investigate the clade of SUS family proteins in which this expansion had occurred, we constructed a phylogenetic tree using characterised SUS protein sequences from *Arabidopsis thaliana*, *Oryza sativa*, *Solanum lycopersicum* and *S. anglica* (Figure 2A).This analysis revealed that the expanded SUS orthogroup in *S. anglica* belonged specifically to the stress-responsive clade I. A subsequent phylogenetic tree was constructed from the clade I SUS family members using sequences from *S. anglica* and closely related plant species (Figure 2B; Supplementary Table S13). This analysis revealed that clade I can be further subdivided into two smaller subclades that we call subclades A and B, and which to our knowledge have not been previously described. Interestingly, the earlier described expansion in the SUS family of *Sporobolus stapfianus* had taken place in subclade A, whereas the *Spartina anglica* expansion had taken place in subclade B, indicating that these family expansion events are unrelated between the two species. Furthermore, within subclade B, a smaller subclade was apparent containing *S. anglica SUS1/4/5* (subclade B.1), which did not follow the taxonomic relationship between the species represented, suggesting that subclade B.1 is the result of specific expansion in *S. anglica* (Figure 2B).

Analysing expression of the *S. anglica SUS* family members in different tissues demonstrated that members of subclade A were most highly expressed in root and rhizome tissues (Figure 2C; Supplementary Table S14). The expression of subclade B genes varied between tissues and, strikingly, members of subclade B.1 showed particularly high expression levels in stem tissue (Figure 2C). Indeed, *SaSUS4* ranked as the 50th most highly expressed gene in stem tissue (Supplementary Table S10). It is not clear if members of these subclades display functional differences from one another, however it is plausible that they are not functionally equivalent given their more distant phylogenetic relationship.

Interestingly, the SUS family appears to have undergone expansion in *Sporobolus stapfianus* as well (Figure 2B; Supplementary Table S13), which also demonstrates considerable abiotic stress tolerance, particularly towards desiccation stress (Chávez Montes et al., 2022). However, this expansion in *S. stapfianus* was to a lesser extent than

in *Spartina anglica* (gain of 3 copies in *Sporobolus stapfianus* vs. 5 copies in *Spartina anglica*). The SUS family also demonstrates contraction in *S. pyramidalis*, which is more sensitive to desiccation than *S. stapfianus* (Figure 2B; Supplementary Table S13). Together, this suggests that the *SUS* gene family is likely central to the physiological stress responses of *S. anglica* growing in the saltmarsh.

The heat shock protein 70 (HSP70) family shows an expansion in *S. anglica*

Another orthogroup that had significantly expanded in *S. anglica* encoded HSP70 family genes ($p = 0.01$), having gained 6 members relative to closely related species (Supplementary Table S11). HSP70 family proteins are ATP-dependent chaperones, implicated in protein folding, temperature response, and response to metals (Usman et al., 2017; Jiang et al., 2020). HSP70 proteins thus have a well-established role in abiotic stress tolerance in plants. In *S. anglica*, the *HSP70* gene family demonstrated an expansion to 18 members (Supplementary Table S11). A phylogenetic tree was constructed with HSP70 protein sequences using the same species as for the SUS family tree. Whilst having more members, the HSP70 family demonstrated lower diversity across all species examined when compared to the SUS family (Supplementary Figure S6; Supplementary Table S15). Additionally, in all cases in *S. anglica*, *HSP70* family members demonstrated the greatest expression levels in root and rhizome tissues (Figure 2D; Supplementary Table S16). Indeed, five of the top ten most highly expressed genes in the rhizome belonged to the expanded *HSP70* family (Figure 2E), suggesting that these genes may be key to environmental rhizome physiology.

***S. anglica* has many ascorbate peroxidases for detoxifying hydrogen peroxide**

Detoxification of reactive oxygen species (ROS), such as H_2O_2, produced in response to abiotic stress is essential for stress tolerance and is achieved by enzymes such as peroxidases. Ascorbate peroxidases (APXs) are type I heme peroxidases that catalyse the H_2O_2-dependent oxidation of ascorbate (Pandey et al., 2017; Lazzarotto et al., 2021; Pang et al., 2023). APXs play an essential role in growth regulation and are found across the plant kingdom. Most plant genomes possess multiple copies of the *APX* gene, with eight *APX* genes identified in *Arabidopsis thaliana, Oryza sativa* and *Zea mays*, and

seven in *Solanum lycopersicum* (Narendra et al., 2006; Najami et al., 2008; Pandey et al., 2017; Ribeiro et al., 2017; Qu et al., 2020). Since APXs play significant roles in plants under a variety of abiotic stresses including saline conditions (Diaz-Vivancos et al., 2013), photo-oxidative damage (Fryer et al., 2003; Pignocchi et al., 2003), temperature fluctuations (Zhang et al., 1997; Sato et al., 2011), drought (Rosa et al., 2010) and metal toxicity (Ekmekçi et al., 2008; Rosa et al., 2010; Pandey et al., 2017), we sought to determine whether *S. anglica* had a large number of APXs.

From the *S. anglica* genome, 10 *APX* gene products were identified (Figure 3A; Supplementary Table S17). Plant APXs can be subdivided into several groups depending on subcellular localisation (Pandey et al., 2017; Leng et al., 2021; Lazzarotto et al., 2021; Pang et al., 2023), and in *S. anglica* eight of the ten identified APXs were predicted to be cytosolic. Only SaAPX8 and SaAPX9 were predicted to be peroxisomal. Interestingly, no chloroplastic APXs were identified. The ten SaAPXs were compared with the APXs from some of the closest relatives of *S. anglica* (*Eragrostis curvula*, *Eleusine coracana* and *Panicum virgatum*, as well as *Z. mays*), and another phylogenetic tree was constructed (Figure 3B; Supplementary Table S18). This tree showed that the most significant difference in phylogenetic relationship occurred between the cytosolic and peroxisomal APXs, with no clear phylogenetic relationship between species for the peroxisomal APXs. However, five of the cytosolic SaAPX proteins were most closely related to a single APX from *E. curvula* and *E. coracana* (EcAPX2 and EIcoAPX4, respectively), suggesting expansion of these genes in *S. anglica*. When looking at expression in *S. anglica*, for the majority of *APX* genes there was little difference in the transcript levels across the different plant tissues (Figure 3C; Supplementary Table S17). However, *APX5* showed higher expression in the roots and rhizome, suggesting that APX5 may play a specific role in decreasing H_2O_2 concentrations in these tissues. This is consistent with *APX* genes being more highly expressed in tissues where oxygen is limited, such as below ground tissues (Blokhina et al., 2001), especially when growing in saltmarsh sediments that are anoxic (Mossman et al., 2020). It is somewhat unexpected that only one out of the ten *APX* genes showed higher expression in the roots and rhizome of *S. anglica*, but it may be that *SaAPX5* expression is triggered at lower local concentrations of H_2O_2 compared with the

other *SaAPX* genes, or that there are temporal differences in expression between *SaAPX5* and the other *SaAPX* genes.

Proteins involved in calcium-mediated stress signalling are also abundant in *S. anglica* and may allow tolerance to salt, heavy metal ions and nutrient stress

Since calcium signalling mediated by calcium-binding Calcineurin B-like proteins (CBLs) and CBL-interacting protein kinases (CIPKs) is essential for responses to abiotic stresses (Batistic and Kudla, 2009), of which there are many in the saltmarsh, we assessed the CBLs and CIPKs in *S. anglica*. CBL and CIPK protein sequences were identified by BLAST searching the *S. anglica* genome using *A. thaliana* and *O. sativa* CBL and CIPK protein sequences as queries (10 CBLs and 26 CIPKs from *A. thaliana*, and 10 CBLs and 33 CIPKs from *O. sativa*). The identified *S. anglica* sequences were then confirmed by determining the presence of key domains: all four EF-hand calcium-binding motifs in CBLs, and the presence of a NAF domain in CIPKs to mediate CBL-CIPK protein interactions (Batistic and Kudla, 2009). In total, 37 CBLs and 69 CIPKs were identified in S. anglica (Supplementary Table S19; Supplementary Table S20). A phylogenetic tree constructed with the identified *S. anglica* CBLs (Figure 4A) showed many CBLs across all groups in *S. anglica* when compared with *Arabidopsis thaliana* and *O. sativa*, but with particular expansion in the Type III transmembrane group. In *Arabidopsis*, AtCBL10 was the only member of this group, but in *S. anglica* the group was expanded to include 11 members (Figure 4A). AtCBL10 is involved in salt tolerance in *A. thaliana*, probably by interaction with AtCIPK24/SOS2 to activate sodium ion pumping into the vacuole to allow compartmentalisation (Kim et al., 2007). Expansion of this Type III CBL group in *S. anglica* may therefore improve tolerance to the highly saline conditions found in the saltmarsh. Analysis of the expression of the *S. anglica CBL* genes showed that *SaCBL4* from this group was highly expressed in all plant tissues (Figure 4B), suggesting that this gene in particular may play an active role in stress signalling and enabling stress tolerance in *S. anglica*. *SaCBL12* and *SaCBL21*, part of the Type I lipid-modified, plasma membrane-targeted CBL group (Edel and Kudla, 2015) were highly expressed in the leaf, rhizome and root. This group contains AtCBL4, also known as SOS3, which is a key protein in the salt tolerance pathway in *A. thaliana* (Ali et al., 2023). It is therefore possible that

SaCBL12 and SaCBL21 also have important roles in the salt tolerance of *S. anglica* growing in the saltmarsh.

The many *S. anglica* CIPKs were also spread across all groups of the protein family, but particularly large numbers were found within the AtCIPK23 group (Edel and Kudla, 2015), where 4 CIPKs in *Arabidopsis* and 6 CIPKs in *O. sativa* were expanded to 24 CIPKs in *S. anglica* (Supplementary Figure S7). AtCIPK23 is a master regulator of ion homeostasis, including heavy metals (Ródenas and Vert, 2021), and is involved in the control of ion channels and transporters for potassium, nitrate, ammonium, manganese, magnesium, cadmium, cobalt, zinc and iron (Xu et al., 2006; Li et al., 2006; Ho et al., 2009; Maierhofer et al., 2014; Tang et al., 2015; Ragel et al., 2015; Tian et al., 2016; Straub et al., 2017; Tang et al., 2020). Since heavy metals are highly abundant in saltmarsh sediments and soils (Payet et al., 2024), and some nutrients such as nitrogen are found at relatively low concentrations, expansion of the AtCIPK23 group in *S. anglica* may be a direct evolutionary response to these pressures to allow increased stress tolerance of the plant to heavy metals and improved regulation of nutrient transport. In support of this, *SaCIPK14*, *SaCIPK34*, *SaCIPK44* and *SaCIPK54* (all encoding AtCIPK23 group members), showed particularly high expression in the root and rhizome tissues (Figure 4C). As roots and rhizomes are important for sensing soil nutrient levels, these four CIPKs may represent important regulators for ion homeostasis signalling pathways in *S. anglica*. Other *S. anglica CIPK* genes that showed high levels of expression in the root included *SaCIPK4* and *SaCIPK23* (from the intronless CIPK group), and *SaCIPK3* and *SaCIPK48* (from the AtCIPK1 group; (Edel and Kudla, 2015). SaCIPK4 and SaCIPK23 are related to AtCIPK18, which has a role in lithium ion homeostasis (Zhang et al., 2020), while AtCIPK1 is involved in the *Arabidopsis* response to nitrate deficiency (Su et al., 2024) and phosphate deficiency (Lu et al., 2020). This could point towards similar roles for SaCIPK3, SaCIPK4, SaCIPK23 and SaCIPK48 in the regulation of ion homeostasis and nutrient stress.

S. anglica produces a range of osmolytes that enhance abiotic stress tolerance in the saltmarsh

S. anglica produces a diverse array of osmolytes at high concentrations, including dimethylsulfoniopropionate (DMSP), glycine betaine (GB) and proline (Cavalieri and Huang, 1981; Payet et al., 2024), some of which are rarely seen at such high levels elsewhere in the plant kingdom, e.g. DMSP (Ausma et al., 2017; Payet et al., 2024). To determine which osmolytes were produced by *S. anglica*, leaf tissue was assayed for a panel of zwitterionic osmolytes known to be produced by other plant species and/or microbes (Supplementary Table S21). Consistent with previous reports, DMSP, GB and proline accumulated at high concentrations (64 μmol g^{-1} DW, 95 μmol g^{-1} DW and 7 μmol g^{-1} DW, respectively; Figure 5A). Surprisingly, we observed proline concentrations far lower (10-14 times) than those of DMSP and GB (Figure 5A), suggesting that proline may represent a less important osmolyte in *S. anglica* than in other plant species.

We then identified genes encoding the complete known synthesis pathways for DMSP, GB and proline in the genome assembly (Supplementary Table S22). For the GB synthesis pathway, homologs of the genes encoding enzymes required for each step of the pathway were identified in *S. anglica* by BLAST using *Oryza sativa* sequences as queries (Luo et al., 2012; Phitaktansakul et al., 2022). In plants, GB is synthesised via choline (Bhuiyan et al., 2007), which itself can act as an osmolyte upon accumulation. However, the choline concentration was 15 times lower than GB in *S. anglica* leaf tissue (Supplementary Table S21), suggesting that choline may be an intermediate of GB synthesis rather than an end-product that is accumulated specifically as an osmolyte in this species. Therefore, the choline synthesis steps were considered to be part of the GB synthesis pathway. Expression analysis showed that genes of the GB and DMSP synthesis pathways were most highly expressed in leaf tissue, whilst genes associated with proline biosynthesis were expressed at similar levels in all five tissue types (Figure 5B). Despite the concentrations of GB detected in leaves being roughly double those of DMSP (Figure 5A), there did not appear to be a corresponding increase in the abundance of GB synthesis genes (Figure 3B). Furthermore, choline monooxygenase (CMO), which bridges the gap between choline biosynthesis and GB biosynthesis, demonstrated very low levels of expression. This suggests either that the GB synthesis enzymes (particularly CMO) display highly favourable kinetics, that substrate availability and/or partitioning

plays a greater role in GB accumulation than DMSP, or that other GB synthesis pathways exist in *S. anglica*. Indeed, other GB synthesis pathways have previously been described in halophytic marine bacteria (Liu et al., 2022), which initiate synthesis from glycine rather than choline and proceed through a sarcosine intermediate. Sarcosine was detected (4 μmol g^{-1} DW) in *S. anglica* leaf extracts (Supplementary Table S21; Figure 5C), however no homologs of bacterial glycine dependent GB synthesis genes (Liu et al., 2022) could be identified in *S. anglica* by a BLAST search.

The zwitterionic alkaloid proline betaine (stachydrine) was also identified (144 μmol g^{-1} DW) in *S. anglica* (Figure 5C). Previous studies have demonstrated that this compound accumulated in plants in response to salinity (Hanson et al., 1994; Trinchant et al., 2004). This, coupled with the chemical properties of proline betaine, support its role as an osmolyte. Interestingly, proline betaine accumulated at 1.5 times the levels of GB, and its total concentration was approximately equivalent to that of GB and DMSP combined (Figure 5C). This suggests that proline betaine may represent one of the main osmolytes in *S. anglica*, which has hitherto been unreported in this genus, or indeed in any grass to the best of our knowledge. To date, no synthesis pathway is known for this compound.

Finally, two further zwitterionic metabolites, ectoine and dimethylsulfoxonium propionate (DMSOP), were also identified in *S. anglica* leaf extracts (56 μmol g^{-1} DW and 10 μmol g^{-1} DW, respectively; Figure 5C). Ectoine has previously been characterised as an osmolyte in halophytic bacteria (Eiberweiser et al., 2015), whilst DMSOP was identified in marine bacteria and microalgae (Thume et al., 2018) and recently shown to act as an osmolyte that mainly functions in oxidative stress protection (Carrión et al., 2023). To the best of our knowledge, neither of these metabolites have thus far been described in a plant.

S. anglica **produces high concentrations of dimethylsulfoniopropionate (DMSP) in part due to a retrotransposon that causes high-level expression of *DMSP-amine oxidase* (*DOX*)**

Spartina plants are some of the highest known producers of the osmolyte DMSP (Payet et al., 2024). We therefore sought to characterise the DMSP synthesis pathway in greater

detail to determine why *S. anglica* produces such high concentrations of DMSP. Components of the DMSP synthesis pathway (*MMT*, *SDC* and *DOX*) were most strongly expressed in leaves (Figure 6A; Supplementary Figure S8), which correlated with leaves accumulating the highest concentrations of DMSP of all tissues in the plant (Payet et al., 2024). We previously showed that the S-methylmethionine decarboxylase (SDC) enzyme is a key determinant of high-level DMSP production in *S. anglica*, but also that expression of *DMSP-amine oxidase* (*DOX*) was far higher (33-84 times) in *S. anglica* than other plant species (Payet et al., 2024). To investigate why this might be, all *DOX* loci were identified in the *S. anglica* genome assembly. Six distinct *DOX* loci were identified in total, and of these, expression of *SaDOX6* exceeded the other five (Figure 6A; Supplementary Table S23). This analysis was extended to *S. alterniflora* (Chen et al., 2024), another *Spartina* species that produces exceptionally high concentrations of DMSP, and identified only four *DOX* loci in the *S. alterniflora* genome. As in *S. anglica*, one locus (*SaltDOX3*) showed considerably higher expression than the others (Figure 6A; Supplementary Table S24). To determine the basis for this novel very high expression profile, 2 kb windows upstream of each of the 6 *DOX* genes in *S. anglica* were retrieved and compared. This revealed a ~2 kb LTR retrotransposon of the gypsy type that was located 651 bp upstream of the transcriptional start site of *SaDOX6* and was completely absent from all the other *DOX* loci. We therefore suspected that this LTR was the cause of high-level expression of the *DOX* gene and subsequently named it LTR^{DOX} (Figure 6B). Analysis of the *S. alterniflora* genome also identified LTR^{DOX} upstream of *SaltDOX3* but not the three other *SaltDOX* genes (Figure 6B).

A BLAST search using LTR^{DOX} as a query sequence was performed to determine other potential insertion sites in the *S. anglica* genome that may have a similar high-level expression profile. This resulted in 518 hits, which were then subset to those within a 1 kb window of a transcriptional start site, yielding only 20 candidates (Figure 6C; Supplementary Table S25). The transcript abundance of each of these 20 genes was examined relative to other members of their orthogroup that did not contain this LTR in their promoters. Of these 20 candidates, none showed high-level expression like *SaDOX6* (Supplementary Table S25). Comparison of the LTR sequences of these candidate gene

promoters with LTR^{DOX} found that the sequences were all highly related, but that LTR^{DOX} occupied its own subclade (Figure 6C). Sequence alignment comparison revealed that the key difference between LTR^{DOX} and the other LTRs was the presence of a unique 55 bp sequence towards the 3' end of LTR^{DOX} that was not seen in any of the other LTRs. Once again, this 55 bp sequence was also found in *SaltDOX3* (Figure 6D). The 55 bp sequence of LTR^{DOX} was analysed for transcription factor binding motifs using PlantPAN 4.0 (Chow et al., 2024) and 49 binding motifs were identified (Supplementary Table S26); primarily, these corresponded to MADS, AT-hook and NF-Y family transcription factor binding sites. To confirm the impact of this region on gene expression, the LTR^{DOX} 55 bp sequence was cloned upstream of the 35S minimal promoter and tested for its ability to drive expression of the firefly luciferase reporter gene. Including the LTR^{DOX} 55 bp sequence increased expression by over 50% relative to a control promoter that lacked this 55 bp sequence (Figure 6E). Overall, this suggests that the LTR^{DOX} 55 bp sequence is important for high-level expression of *SaDOX*.

DISCUSSION

Expansion of gene families can result in neofunctionalization, as the introduction of redundancy can free a gene from the consequences of alteration of function on the original process. This can ultimately result in genes that have either new functions or new regulatory patterns. In *S. anglica*, we observed an expansion in the stress-responsive *SUS* family of sucrose synthases. Concurrent with this, we observed a distinct subclade of SUS proteins (clade B.1) that had seemingly emerged in *S. anglica*, as it was absent in closely related species (Figure 2B). This clade appears to have a different expression pattern from the other *SUS* genes, showing high expression in the stem of *S. anglica*. As *S. anglica* occupies saltmarshes and estuaries, the vascular tissues are routinely perfused with highly saline sea water. One function that *SUS* subclade B.1 may serve is to counterbalance the osmotic potential difference in these tissues to minimise water loss in this way. Clade I SUS family members had also expanded in the desiccation-tolerant *Sporobolus stapfianus* and contracted in the desiccation-sensitive *Sporobolus pyramidalis* (Figure 2B), indicating that this family may be highly relevant to plant water status.

The HSP70 orthogroup had undergone the largest expansion in *S. anglica*. Furthermore, this group contained some of the most highly expressed genes in the rhizomes of *S. anglica* (Figure 2E). Tolerance to high levels of abiotic stresses such as drought, salinity and inundation are central to the persistence of *S. anglica* rhizomes, and this, in turn, is critical for the invasive nature of this species (Castillo et al., 2017). It is possible that expansion of the HSP70 family may therefore represent one of the genetic innovations which underpin the highly invasive nature of *S. anglica* (Borges et al., 2021). The large abundance of genes involved in the detoxification of ROS (Figure 3) and in calcium signalling (Figures 4 and 5) could also contribute to the ability of *S. anglica* to tolerate high levels of abiotic stress in the saltmarsh.

In addition to SUS and HSP70, 147 other orthogroups showed evidence of undergoing expansion (Supplementary Table S11). Given the strong, abiotic selection pressure of saltmarshes, it is reasonable that many of these other expanded families may also convey abiotic stress tolerance. These may therefore represent promising candidate genes for future efforts to bioengineer crops that are more tolerant to environmental stress. Indeed, there are already examples emerging of *S. anglica* and *S. alterniflora* genes being used to enhance abiotic stress tolerance in plants (Yang et al., 2023; Payet et al., 2024; Wei et al., 2024). The transcriptome of *S. anglica* displayed many GO process terms related to water status of the plant (Supplementary Figure S4) and it is interesting to note that, as a C4 species, *Spartina anglica* likely has improved water use efficiency over other saltmarsh C3 plants (Monson et al., 2025). Moreover, nitrogen use efficiency is also improved in C4 plants over C3 plants (Monson et al., 2025). It is therefore conceivable that the evolutionary links between C4 photosynthesis, water use efficiency and nitrogen use efficiency may have contributed to the ability of *S. anglica* to colonise the waterlogged and nitrogen-poor saltmarsh environment so effectively since its relatively recent formation in ~1870. The availability of *S. anglica* genomic resources here may now also help identify the molecular mechanisms underpinning other interesting traits found in wetland plant species. For example, comparison between *S. anglica* and *S. alterniflora* genomes may identify the genetic basis underlying formation of lysigenous aerenchyma,

which is reported to be sensitive to flooding in *S. alterniflora* but unaffected by flooding in *S. anglica* (Maricle and Lee, 2002). *S. anglica* also displayed a rich osmolyte profile, which contained zwitterionic compounds that have not to date been described in plants (e.g. ectoine and DMSOP) and compounds for which biosynthesis enzymes are not currently known in plants (e.g. ectoine, DMSOP and stachydrine) (Figure 5C; Supplementary Table S21). Given the high concentrations of these compounds accumulated in *S. anglica*, the genomic data presented here also offer a useful resource for novel pathway discovery. Whether these different osmolytes play different roles or accumulate at different concentrations in various tissue types also remains to be explored in the future.

An outstanding question in the evolution of high-level DMSP synthesis in *Spartina* species was how the elevated *DOX* expression seen in these species evolved (Payet et al., 2024). Our genome-wide analysis demonstrated that in both *S. anglica* and *S. alterniflora*, most *DOX* transcripts come from one locus (*SaDOX6* and *SaltDOX3*, respectively). In both cases, an LTR gypsy family retrotransposon was situated within 1 kb upstream of the transcriptional start site, which we call LTR^{DOX} (Figure 6B). This LTR contained a 55 bp region which was unique to the *DOX* loci (Figure 6D) and could increase gene expression by over 50% (Figure 6E). It is possible that a random 55 bp sequence could be sufficient to activate expression of the 35S minimal promoter, so further work, including mutational approaches, would be interesting to confirm the role of LTR^{DOX} in increasing gene expression. However, in our previous work, we identified an ethylene-responsive transcription factor (EREBP1) that may regulate DMSP synthesis in *S. anglica* (Payet et al., 2024) and it is interesting to note that an ethylene-responsive AP2/ERF binding motif was identified in the unique 55 bp region of LTR^{DOX} (Supplementary Table S26). Whether this AP2/ERF binding motif is regulated by EREBP1 would be interesting to test in the future. Overall, the example of LTR^{DOX} represents a good demonstration of how movement of transposable elements can influence gene expression and support the evolution of new traits. Future work to further mine the *S. anglica* genomic and transcriptomic responses here will no doubt yield additional fascinating insights into the evolution of this species and the novel adaptations present in the *Spartina* lineage.

ACKNOWLEDGEMENTS

We thank the National Trust for allowing us to sample *Spartina anglica* from Stiffkey saltmarsh and NERC Environmental Omics Facility for access to the facilities in Sheffield and Liverpool.

AUTHOR CONTRIBUTIONS

RDP and JBM conceived and designed the experiments. RDP and SMTK performed most of the bioinformatic analyses. AMF and AMER performed the analysis of APXs and CBL-CIPKs, respectively. LJB performed RNA extractions for transcriptomics. MEG built constructs to test LTR^{DOX} function in *N. benthamiana*. CO calculated bootstrap values and constructed phylogenetic trees. LSK performed extraction of high molecular weight DNA for genomic sequencing. XL performed initial genome assembly analysis. MA and GP performed osmolyte analysis. AJD oversaw *S. anglica* sampling and provided ecological insight. JBM performed the LTR^{DOX} promoter testing experiment. JDT and JBM supervised the project. RDP and JBM interpreted the data and wrote the paper, with contributions from SMTK, AMF and AMER. All authors reviewed and approved the manuscript.

CONFLLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

FUNDING STATEMENT

This work was supported by a NERC Environmental Omics Facility grant (NEOF1369). RDP, LJB, AMF, AJD, JDT and JBM were funded by the Natural Environment Research Council (NERC; NE/V000756/1). RDP, LJB, JDT and JBM were also funded by the Biotechnology and Biological Sciences Research Council (BB/X005968/1). SMTK and AMER were funded by PhD studentships from NERC via the ARIES Doctoral Training Partnership (NE/S007334/1) and BBSRC via the Norwich Research Park Doctoral Training Partnership (NRPDTP) (BB/T008717/1), respectively. MEG was supported by an Undergraduate Summer Internship Scheme bursary via the NRPDTP. GP and MA were

funded by the German Research Foundation (DFG) Project-ID 239748522 (SFB 1127 ChemBioSys).

DATA AVAILABILITY STATEMENT

The genome assembly and annotation files generated in this study have been deposited in the Zenodo repository at https://doi.org/10.5281/zenodo.21354741. The transcriptomic data have been deposited in the NCBI database under BioProject accession number PRJNA1495156.

REFERENCES

Ainouche ML, Baumel A, Salmon A, Yannic G. 2004a. Hybridization, polyploidy and speciation in *Spartina* (Poaceae). *New Phytologist* **161**:165–172.

Ainouche ML, Baumel A, Salmon A. 2004b. *Spartina anglica* C. E. Hubbard: A natural model system for analysing early evolutionary changes that affect allopolyploid genomes. *Biological Journal of the Linnean Society* **82**:475–484.

Alexa A, Rahnenführer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**:1600–1607.

Ali A, Petrov V, Yun D-J, Gechev T. 2023. Revisiting plant salt tolerance: Novel components of the SOS pathway. *Trends in Plant Science* **28**:1060–1069.

Anburajan L, Meena B, Vinithkumar NV, Dharani G. 2021. Molecular characterization of glycine betaine biosynthesis genes from deep sea halophilic bacteria, *Bacillus atrophaeus* NIOT-DSB21. *Ecological Genetics and Genomics* **18**:100080.

Ausma T, Kebert M, Stefels J, De Kok LJ. 2017. DMSP: Occurrence in plants and response to salinity in *Zea mays*. In: De Kok LJ, Hawkesford MJ, Haneklaus SH, Schnug E, eds. Sulfur Metabolism in Higher Plants - Fundamental, Environmental and Agricultural Aspects. Cham: Springer International Publishing, 87–91.

Batistic O, Kudla J. 2009. Plant calcineurin B-like proteins and their interacting protein kinases. *Biochimica et Biophysica Acta* **1793**:985–992.

Baumel A, Ainouche M, Kalendar R, Schulman AH. 2002. Retrotransposons and genomic stability in populations of the young allopolyploid species *Spartina anglica* C.E. Hubbard (Poaceae). *Molecular Biology and Evolution* **19**:1218–1227.

Beeftink WG, Nieuwenhuize J, Stoeppler M, Mohl C. 1982. Heavy-metal accumulation in salt marshes from the Western and Eastern Scheldt. *Science of The Total Environment* **25**:199–223.

Bhuiyan NH, Hamada A, Yamada N, Rai V, Hibino T, Takabe T. 2007. Regulation of betaine synthesis by precursor supply and choline monooxygenase expression in *Amaranthus tricolor*. *Journal of Experimental Botany* **58**:4203–4212.

Blázquez MA. 2024. Polyamines: Their role in plant development and stress. *Annual Review of Plant Biology* **75**:95–117.

Blokhina OB, Chirkova TV, Fagerstedt KV. 2001. Anoxic stress leads to hydrogen peroxide formation in plant cells. *Journal of Experimental Botany* **52**:1179–1190.

Borges FO, Santos CP, Paula JR, *et al.* 2021. Invasion and extirpation potential of native and invasive *Spartina* species under climate change. *Frontiers in Marine Science* **8**.

Bortolus A, Adam P, Adams JB, *et al.* 2019. Supporting *Spartina*: Interdisciplinary perspective shows *Spartina* as a distinct solid genus. *Ecology* **100**:e02863.

Burden A, Garbutt A, Evans CD. 2019. Effect of restoration on saltmarsh carbon accumulation in Eastern England. *Biology Letters* **15**:20180773.

Burke DJ, Weis JS, Weis P. 2000. Release of metals by the leaves of the salt marsh grasses *Spartina alterniflora* and *Phragmites australis*. *Estuarine, Coastal and Shelf Science* **51**:153–159.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**:188–196.

Card DC, Adams RH, Schield DR, *et al.* 2019. Genomic basis of convergent island phenotypes in Boa constrictors. *Genome Biology and Evolution* **11**:3123–3143.

Carrión O, Li C-Y, Peng M, *et al.* 2023. DMSOP-cleaving enzymes are diverse and widely distributed in marine microorganisms. *Nature Microbiology* **8**:2326–2337.

Castillo J, Leira-Doce P, Figueroa E. 2017. Biomass and clonal architecture of the cordgrass *Spartina patens* (Poaceae) as an invasive species in two contrasted coastal habitats on the Atlantic coast of the Iberian Peninsula. *Plant Ecology and Evolution* **150**:129–138.

Cavalieri AJ, Huang AHC. 1981. Accumulation of proline and glycinebetaine in *Spartina alterniflora* Loisel. in response to NaCl and nitrogen in the marsh. *Oecologia* **49**:224–228.

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. 2020. BlobToolKit – Interactive Quality Assessment of Genome Assemblies. *G3 Genes|Genomes|Genetics* **10**:1361–1374.

Chávez Montes RA, Haber A, Pardo J, *et al.* 2022. A comparative genomics examination of desiccation tolerance and sensitivity in two sister grass species. *Proceedings of the National Academy of Sciences* **119**:e2118886119.

Chen S, Du T, Huang Z, *et al.* 2024. The *Spartina alterniflora* genome sequence provides insights into the salt-tolerance mechanisms of exo-recretohalophytes. *Plant Biotechnology Journal* **22**:2558–2574.

Chi YH, Koo SS, Oh HT, *et al.* 2019. The physiological functions of universal stress proteins and their molecular mechanism to protect plants from environmental stresses. *Frontiers in Plant Science* **10**.

Chow C-N, Yang C-W, Wu N-Y, Wang H-T, Tseng K-C, Chiu Y-H, Lee T-Y, Chang W-C. 2024. PlantPAN 4.0: Updated database for identifying conserved non-coding sequences and exploring

dynamic transcriptional regulation in plant promoters. *Nucleic Acids Research* **52**:D1569–D1578.

Curson ARJ, Todd JD, Sullivan MJ, Johnston AWB. 2011. Catabolism of dimethylsulphoniopropionate: Microorganisms, enzymes and genes. *Nature Reviews Microbiology* **9**:849–859.

De Bie T, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **22**:1269–1271.

Demuth JP, Hahn MW. 2009. The life and death of gene families. *Bioessays* **31**:29–39.

Diaz-Vivancos P, Faize M, Barba-Espin G, Faize L, Petri C, Hernández JA, Burgos L. 2013. Ectopic expression of cytosolic superoxide dismutase and ascorbate peroxidase leads to salt stress tolerance in transgenic plums. *Plant Biotechnology Journal* **11**:976–985.

Edel KH, Kudla J. 2015. Increasing complexity and versatility: How the calcium signaling toolkit was shaped during plant land colonization. *Cell Calcium* **57**:231–246.

Eiberweiser A, Nazet A, Kruchinin SE, Fedotova MV, Buchner R. 2015. Hydration and ion binding of the osmolyte ectoine. *Journal of Physical Chemistry B* **119**:15203–15211.

Ekmekçi Y, Tanyolaç D, Ayhan B. 2008. Effects of cadmium on antioxidant enzyme and photosynthetic activities in leaves of two maize cultivars. *Journal of Plant Physiology* **165**:600–611.

Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**:238.

Feike D, Korolev AV, Soumpourou E, *et al.* 2019. Characterizing standard genetic parts and establishing common principles for engineering legume and cereal roots. *Plant Biotechnology Journal* **17**:2234–2245.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**:9451–9457.

Fryer MJ, Ball L, Oxborough K, Karpinski S, Mullineaux PM, Baker NR. 2003. Control of *Ascorbate Peroxidase 2* expression by hydrogen peroxide and leaf water status during excess light stress reveals a functional organisation of *Arabidopsis* leaves. *The Plant Journal* **33**:691–705.

Global Invasive Species Database. https://www.iucngisd.org/gisd/100_worst.php. Accessed March 17, 2025.

Grabherr MG, Haas BJ, Yassour M, *et al.* 2011. Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* **29**:644.

Granse D, Titschack J, Ainouche M, Jensen K, Koop-Jakobsen K. 2022. Subsurface aeration of tidal wetland soils: Root-system structure and aerenchyma connectivity in *Spartina* (Poaceae). *Science of the Total Environment* **802**:149771.

Guo L-M, Li J, He J, Liu H, Zhang H-M. 2020. A class I cytosolic HSP20 of rice enhances heat and salt tolerance in different organisms. *Scientific Reports* **10**:1383.

Gurrieri L, Merico M, Trost P, Forlani G, Sparla F. 2020. Impact of drought on soluble sugars and free proline content in selected *Arabidopsis* mutants. *Biology* **9**:367.

Hanson AD, Rathinasabapathi B, Rivoal J, Burnet M, Dillon MO, Gage DA. 1994. Osmoprotective compounds in the Plumbaginaceae: A natural experiment in metabolic engineering of stress tolerance. *Proceedings of the National Academy of Sciences* **91**:306–310.

Hao Y, Wang X-F, Guo Y, *et al.* 2024. Genomic and phenotypic signatures provide insights into the wide adaptation of a global plant invader. *Plant Communications* **5**:100820.

Ho C-H, Lin S-H, Hu H-C, Tsay Y-F. 2009. CHL1 functions as a nitrate sensor in plants. *Cell* **138**:1184–1194.

Hou W, Liang S, Hu Z, Sun Z, Ma Q, Deng Z, Zhang R. 2024. Sea-level rise versus salt marsh colonization: The adversarial game of self-organized elevation maintenance in tidal marsh. *Science of The Total Environment* **954**:176554.

Huerta-Cepas J, Szklarczyk D, Heller D, *et al.* 2019. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**:D309–D314.

Ji X-R, Yu Y-H, Ni P-Y, Zhang G-H, Guo D-L. 2019. Genome-wide identification of small heat-shock protein (HSP20) gene family in grape and expression profile during berry development. *BMC Plant Biology* **19**:433.

Jiang C, Bi Y, Zhang R, Feng S. 2020. Expression of *RcHSP70*, heat shock protein 70 gene from Chinese rose, enhances host resistance to abiotic stresses. *Scientific Reports* **10**:2445.

Jones P, Binns D, Chang H-Y, *et al.* 2014. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**:1236.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**:587–589.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**:772–780.

Kim B-G, Waadt R, Cheong YH, Pandey GK, Dominguez-Solis JR, Schültke S, Lee SC, Kudla J, Luan S. 2007. The calcium sensor CBL10 mediates salt tolerance by regulating ion homeostasis in Arabidopsis. *The Plant Journal* **52**:473–484.

Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* **5**:59.

Lazzarotto F, Menguer PK, Del-Bem LE, Zámocký M, Margis-Pinheiro M. 2021. Ascorbate peroxidase neofunctionalization at the origin of APX-R and APX-L: Evidence from basal Archaeplastida. *Antioxidants* **10**:597.

Leng X, Wang H, Zhang S, Qu C, Yang C, Xu Z, Liu G. 2021. Identification and characterization of the *APX* gene family and its expression pattern under phytohormone treatment and abiotic stress in *Populus trichocarpa*. *Genes* **12**:334.

Li L, Kim B-G, Cheong YH, Pandey GK, Luan S. 2006. A Ca^{2+} signaling pathway regulates a K^+ channel for low-K response in *Arabidopsis*. *Proceedings of the National Academy of Sciences* **103**:12625–12630.

Lian X, Wang Q, Li T, *et al.* 2022. Phylogenetic and transcriptional analyses of the *HSP20* gene family in peach revealed that *PpHSP20-32* is involved in plant height and heat tolerance. *International Journal of Molecular Sciences* **23**:10849.

Liu M, Liu H, Mei F, Yang N, Zhao D, Ai G, Xiang H, Zheng Y. 2022. Identification of the biosynthetic pathway of glycine betaine that is responsible for salinity tolerance in halophilic *Thioalkalivibrio versutus* D301. *Frontiers in Microbiology* **13**.

Liu L, Li X, Wang C, Ni Y, Liu X. 2023. The role of chloride channels in plant responses to NaCl. *International Journal of Molecular Sciences* **25**:19.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550.

Lu X, Li X, Xie D, Jiang C, Wang C, Li L, Zhang Y, Tian H, Gao H, Wang C. 2020. The Ca^{2+}-regulated protein kinase CIPK1 integrates plant responses to phosphate deficiency in *Arabidopsis thaliana*. *Plant Biology* **22**:753–760.

Lukashin AV, Borodovsky M. 1998. GeneMark.hmm: New solutions for gene finding. *Nucleic Acids Research* **26**:1107–1115.

Luo D, Niu X, Yu J, Yan J, Gou X, Lu B-R, Liu Y. 2012. Rice choline monooxygenase (OsCMO) protein functions in enhancing glycine betaine biosynthesis in transgenic tobacco but does not accumulate in rice (*Oryza sativa* L. ssp. *japonica*). *Plant Cell Reports* **31**:1625–1635.

Maierhofer T, Diekmann M, Offenborn JN, *et al.* 2014. Site- and kinase-specific phosphorylation-mediated activation of SLAC1, a guard cell anion channel stimulated by abscisic acid. *Science Signaling* **7**:ra86.

Maricle BR, Lee RW. 2002. Aerenchyma development and oxygen transport in the estuarine cordgrasses *Spartina alterniflora* and *S. anglica*. *Aquatic Botany* **74**:109–120.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**:10–12.

Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M. 2005. The NAC transcription factors NST1 and NST2 of *Arabidopsis* regulate secondary wall thickenings and are required for anther dehiscence. *The Plant Cell* **17**:2993.

Monson RK, Li S, Ainsworth EA, *et al.* 2025. C4 photosynthesis, trait spectra, and the fast-efficient phenotype. *New Phytologist* **246**:879–893.

Mossman HL, Grant A, Davy AJ. 2020. Manipulating saltmarsh microtopography modulates the effects of elevation on sediment redox potential and halophyte distribution. *Journal of Ecology* **108**:94–106.

Najami N, Janda T, Barriah W, Kayam G, Tal M, Guy M, Volokita M. 2008. Ascorbate peroxidase gene family in tomato: its identification and characterization. *Molecular Genetics and Genomics* **279**:171–182.

Nansheng C. 2004. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* **Chapter 4**.

Narendra S, Venkataramani S, Shen G, Wang J, Pasapula V, Lin Y, Kornyeyev D, Holaday AS, Zhang H. 2006. The *Arabidopsis* ascorbate peroxidase 3 is a peroxisomal membrane-bound antioxidant enzyme and is dispensable for *Arabidopsis* growth and development. *Journal of Experimental Botany* **57**:3033–3042.

Pandey S, Fartyal D, Agarwal A, Shukla T, James D, Kaul T, Negi YK, Arora S, Reddy MK. 2017. Abiotic stress tolerance in plants: Myriad roles of ascorbate peroxidase. *Frontiers in Plant Science* **8**.

Pang X, Chen J, Xu Y, Liu J, Zhong Y, Wang L, Zheng J, Wan H. 2023. Genome-wide characterization of ascorbate peroxidase gene family in pepper (*Capsicum annuum* L.) in response to multiple abiotic stresses. *Frontiers in Plant Science* **14**.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**:417–419.

Payet RD, Bilham LJ, Kabir SMT, *et al.* 2024. Elucidation of *Spartina* dimethylsulfoniopropionate synthesis genes enables engineering of stress tolerant plants. *Nature Communications* **15**:8568.

Phitaktansakul R, Kim K-W, Aung KM, *et al.* 2022. Multi-omics analysis reveals the genetic basis of rice fragrance mediated by *betaine aldehyde dehydrogenase 2*. *Journal of Advanced Research* **42**:303–314.

Pignocchi C, Fletcher JM, Wilkinson JE, Barnes JD, Foyer CH. 2003. The function of ascorbate oxidase in tobacco. *Journal of Plant Physiology* **132**:1631–1641.

Qi X, Chmura GL. 2023. Invasive *Spartina alterniflora* marshes in China: a blue carbon sink at the expense of other ecosystem services. *Frontiers in Ecology and the Environment* **21**:182–190.

Qu C, Wang L, Zhao Y, Liu C. 2020. Molecular evolution of maize ascorbate peroxidase genes and their functional divergence. *Genes* **11**:1204.

Ragel P, Ródenas R, García-Martín E, *et al.* 2015. The CBL-interacting protein kinase CIPK23 regulates HAK5-mediated high-affinity K^+ uptake in Arabidopsis roots. *Plant Physiol* **169**:2863–2873.

Renny-Byfield S, Ainouche M, Leitch IJ, Lim KY, Le Comber SC, Leitch AR. 2010. Flow cytometry and GISH reveal mixed ploidy populations and *Spartina* nonaploids with genomes of *S. alterniflora* and *S. maritima* origin. *Annals of Botany* **105**:527–533.

Ribeiro CW, Korbes AP, Garighan JA, Jardim-Messeder D, Carvalho FEL, Sousa RHV, Caverzan A, Teixeira FK, Silveira JAG, Margis-Pinheiro M. 2017. Rice peroxisomal ascorbate peroxidase knockdown affects ROS signaling and triggers early leaf senescence. *Plant Science* **263**:55–65.

Ródenas R, Vert G. 2021. Regulation of root nutrient transporters by CIPK23: “One kinase to rule them all.” *Plant Cell Physiology* **62**:553–563.

Rosa SB, Caverzan A, Teixeira FK, Lazzarotto F, Silveira JAG, Ferreira-Silva SL, Abreu-Neto J, Margis R, Margis-Pinheiro M. 2010. Cytosolic APX knockdown indicates an ambiguous redox responses in rice. *Phytochemistry* **71**:548–558.

Rozema J, Dorel F, Janissen R, Lenssen G, Broekman R, Arp W, Drake BG. 1991. Effect of elevated atmospheric CO_2 on growth, photosynthesis and water relations of salt marsh grass species. *Aquatic Botany* **39**:45–55.

Salmon A, Hao Y, Milin M, *et al.* 2025. On the way to diploidization and unexpected ploidy in the grass *Sporobolus* section *Spartina* mesopolyploids. *Nature Communications* **16**:1997.

Sato Y, Masuta Y, Saito K, Murayama S, Ozawa K. 2011. Enhanced chilling tolerance at the booting stage in rice by transgenic overexpression of the ascorbate peroxidase gene, *OsAPXa*. *Plant Cell Reports* **30**:399–406.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research* **34**:W435–W439.

Stein O, Granot D. 2019. An overview of sucrose synthases in plants. *Frontiers in Plant Science* **10**.

Straub T, Ludewig U, Neuhäuser B. 2017. The kinase CIPK23 inhibits ammonium transport in *Arabidopsis thaliana*. *The Plant Cell* **29**:409–422.

Su H, Wang Q, Wang L, Cui J. 2024. The Ca^{2+}-regulated protein kinase CIPK1 modulates plant response to nitrate deficiency in *Arabidopsis*. *Genes* **15**:1235.

Tang R-J, Zhao F-G, Garcia VJ, Kleist TJ, Yang L, Zhang H-X, Luan S. 2015. Tonoplast CBL–CIPK calcium signaling network regulates magnesium homeostasis in Arabidopsis. *Proceedings of the National Academy of Sciences* **112**:3134–3139.

Tang R-J, Zhao F-G, Yang Y, Wang C, Li K, Kleist TJ, Lemaux PG, Luan S. 2020. A calcium signalling network activates vacuolar K^+ remobilization to enable plant adaptation to low-K environments. *Nature Plants* **6**:384–393.

Tansley C, Houghton J, Rose AME, Witek B, Payet RD, Wu T, Miller JB. 2023. CIPK-B is essential for salt stress signalling in *Marchantia polymorpha*. *New Phytologist* **237**:2210–2223.

The UniProt Consortium 2025. UniProt: The universal protein knowledgebase in 2025. *Nucleic Acids Research* **53**:D609–D617.

Thume K, Gebser B, Chen L, Meyer N, Kieber DJ, Pohnert G. 2018. The metabolite dimethylsulfoxonium propionate extends the marine organosulfur cycle. *Nature* **563**:412–415.

Tian Q, Zhang X, Yang A, Wang T, Zhang W-H. 2016. CIPK23 is involved in iron acquisition of Arabidopsis by affecting ferric chelate reductase activity. *Plant Science* **246**:70–79.

Trinchant J-C, Boscari A, Spennato G, Van de Sype G, Le Rudulier D. 2004. Proline betaine accumulation and metabolism in alfalfa plants under sodium chloride stress. Exploring its compartmentalization in nodules. *Journal of Plant Physiology* **135**:1583–1594.

Usman MG, Rafii MY, Martini MY, Yusuff OA, Ismail MR, Miah G. 2017. Molecular analysis of Hsp70 mechanisms in plants and their function in response to stress. *Biotechnology and Genetic Engineering Reviews* **33**:26–39.

Vasquez EA, Glenn EP, Guntenspergen GR, Brown JJ, Nelson SG. 2006. Salt tolerance and osmotic adjustment of *Spartina alterniflora* (Poaceae) and the invasive M haplotype of *Phragmites australis* (Poaceae) along a salinity gradient. *American Journal of Botany* **93**:1784–1790.

Veldhuis ER, Schrama M, Staal M, Elzenga JTM. 2019. Plant stress-tolerance traits predict salt marsh vegetation patterning. *Frontiers in Marine Science* **5**.

Wang Y, Tang H, DeBarry JD, *et al.* 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* **40**:e49.

Wang H, Sui X, Guo J, Wang Z, Cheng J, Ma S, Li X, Zhang Z. 2014. Antisense suppression of cucumber (*Cucumis sativus* L.) sucrose synthase 3 (*CsSUS3*) reduces hypoxic stress tolerance. *Plant, Cell & Environment* **37**:795–810.

Wei X, Xu L, Dong S, *et al.* 2024. *SaTDT* enhanced plant tolerance to NaCl stress by modulating the levels of malic acid and citric acid in cells. *Plant Molecular Biology* **115**:4.

Wong TKF, Ly-Trong N, Ren H, *et al.* 2025. IQ-TREE 3: Phylogenomic inference software using complex evolutionary models. *Molecular Biology and Evolution* **43**:msag117.

Xu J, Li H-D, Chen L-Q, Wang Y, Liu L-L, He L, Wu W-H. 2006. A protein kinase, interacting with two calcineurin B-like proteins, regulates K^+ transporter AKT1 in *Arabidopsis*. *Cell* **125**:1347–1360.

Xu X, Yang Y, Liu C, Sun Y, Zhang T, Hou M, Huang S, Yuan H. 2019. The evolutionary history of the sucrose synthase gene family in higher plants. *BMC Plant Biology* **19**:566.

Yang Y, Li Y, Chen Q, Sun Y, Lu Z. 2019. WGDdetector: a pipeline for detecting whole genome duplication events using the genome or transcriptome annotations. *BMC Bioinformatics* **20**:75.

Yang M, Chen S, Huang Z, *et al.* 2023. Deep learning-enabled discovery and characterization of genes in *Spartina alterniflora*. *The Plant Journal* **116**:690–705.

Zhang H, Wang J, Nickel U, Allen RD, Goodman HM. 1997. Cloning and expression of an Arabidopsis gene encoding a putative peroxisomal ascorbate peroxidase. *Plant Molecular Biology* **34**:967–971.

Zhang X, Li Z, Li X, Xu Y, Xie H, Qiu Q-S. 2020. CBL3 and CIPK18 are required for the function of NHX5 and NHX6 in mediating Li^+ homeostasis in Arabidopsis. *Journal of Plant Physiology* **255**:153295.

TABLES

Table 1: Genome assembly and annotation statistics. N50 and N90 represent the length of the shortest contig for which all contigs of equal or greater length cover for at least 50% and 90% of the assembly, respectively. GC refers to the guanine-cytosine content.

Assembly feature	Value
Assembled genome size (Gb)	2.97
Number of contigs	497
Total length of contigs (bp)	2,978,325,964
N50 of contigs (bp)	48,744,091
N90 of contigs (bp)	28,480,330
GC content (%)	45.18
Transposable elements (%)	71.3
Masked repeat sequence length (bp)	2,140,632,283
Predicted protein-coding genes	91,935
Complete BUSCOs (%)	92.8

FIGURE LEGENDS

Figure 1: Genomic features, phylogeny and synteny of *Spartina anglica*

A. Circular representation of the *Spartina anglica* genome assembly showing, from outer to inner rings, mapping read counts (100 kb window size), gene density (100 kb window size), GC content (10 kb window size), contig layout, and syntenic gene relationships. Different colours represent links among different contigs. **B.** Phylogenetic tree showing the placement of *S. anglica* within the Commelinids clade. **C.** Microsynteny of *Spartina anglica* with other Chloridoideae grass species. The syntenic regions illustrate ploidy levels, with *Oropetium thomaeum* as a diploid (middle), *Sporobolus stapfianus* as a tetraploid (top), and *S. pyramidalis* as a hexaploidy (top). *S. anglica* is shown as a dodecaploid (bottom). Grey lines indicate collinear connections, while green and blue regions represent gene clusters.

Alt text: Diagrams labelled A to C. A is a circular representation of the *Spartina anglica* genome. B is a phylogenetic tree showing the placement of *S. anglica* within the Commelinids clade. C illustrates the syntenic regions and ploidy levels between *S. anglica* and related species.

Figure 2: SUS and HSP70 protein families associated with stress tolerance are expanded in *Spartina anglica*

A. Phylogenetic tree constructed using SUS protein sequences from *Oryza sativa* (Os), *Arabidopsis thaliana* (At), *Solanum lycopersicum* (Sl), and expanded SUS Clade I orthogroup sequences from *Spartina anglica* (Sa). Colours indicate different clades and dots indicate *S. anglica* sequences. For visual clarity, Clade II and Clade III orthogroup sequences from *S. anglica* are not shown. **B.** Phylogenetic tree constructed using SUS orthogroup sequences from related monocotyledonous plants (*Eragrostis curvula* (Ec), *Musa accuminata* (Ma), *Spartina anglica* (Sa), *Sporobolus pyramidalis* (Sp), *Sporobolus stapfianus* (Ss), and *Zea mays* (Zm)) demonstrating that the expansion in *S. anglica* has occurred in subclade B and that a new subclade (subclade B.1) has emerged. Dots indicate *S. anglica* sequences and red text indicates subclade B.1. **C.** Heatmap showing expression (DESeq2 normalised read numbers) of *S. anglica SUS* orthogroup family members in different tissue types. Orange and blue text indicate genes from subclades A and B, respectively. **D.** Heatmap showing expression (DESeq2 normalised read numbers) of *S. anglica HSP70* genes in different tissue types. **E.** Mean read numbers of top 20 most abundant genes in the transcriptome of *S. anglica* rhizome tissue, demonstrating that the expanded *HSP70* family members occupy six of these positions. Bar colours indicate the family to which the gene belongs.

Alt text: Diagrams and graphs labelled A to E. A and B are phylogenetic trees showing SUS sequences from different plant species. C and D are a heatmap showing expression of *Spartina anglica SUS* and *HSP70* genes in different tissues. E is a graph showing *HSP70* transcript abundance in rhizome tissue.

Figure 3: *Spartina anglica* has many stress-responsive ascorbate peroxidase (APX) proteins

A. Phylogenetic tree constructed using APX protein sequences from *Oryza sativa* (Os), *Arabidopsis thaliana* (At), *Zea mays* (Zm) and *Spartina anglica* (Sa). Colours indicate different clades, which divide according to subcellular protein localisation and dots indicate *S. anglica* sequences. **B.** Phylogenetic tree constructed using APX protein sequences from related monocotyledonous plants (*Eragrostis curvula* (Ec), *Eleusine coracana* (Elco), *Musa accuminata* (Ma), *Spartina anglica* (Sa), *Panicum virgatum* (Pv), and *Zea mays* (Zm)). Dots indicate *S. anglica* sequences. **C.** Heatmap showing expression (DESeq2 normalised read numbers) of *S. anglica APX* genes in different tissue types. Orange and blue text indicate genes from the peroxisomal and cytosolic subclades, respectively.

Alt text: Diagrams labelled A to C. A and B are phylogenetic trees showing APX sequences from different plant species. C is a heatmap showing expression of *Spartina anglica APX* genes in different tissues.

Figure 4: Calcineurin B-like (CBL) and CBL-interacting protein kinase (CIPK) proteins involved in calcium-mediated stress signalling are abundant in *Spartina anglica*

A. Phylogenetic tree constructed using CBL protein sequences from *Oryza sativa* (OsCBL), *Arabidopsis thaliana* (AtCBL) and *Spartina anglica* (SaCBL). Colours indicate different clades, which divide according to subcellular protein localisation based on Edel and Kudla (2015): Type I lipid-modified, plasma membrane-targeted; Type II lipid-modified, tonoplast-targeted; Type III transmembrane domain. Dots indicate *S. anglica* sequences. **B.** Heatmap showing expression (DESeq2 normalised read numbers) of *S. anglica CBL* genes in different tissue types. Orange, blue and red text indicate genes from the different clades shown in panel A. **C.** Heatmap showing expression (DESeq2 normalised read numbers) of *S. anglica CIPK* genes in different tissue types.

Alt text: Diagrams labelled A to C. A is a phylogenetic tree showing CBL sequences from different plant species. B and C are heatmaps showing expression of *Spartina anglica* *CBL* and *CIPK* genes in different tissues.

Figure 5: *Spartina anglica* produces a diverse range of osmolytes

A. Abundance of DMSP, glycine betaine and proline in *Spartina anglica* leaf tissue. **B.** Heatmap showing expression (DESeq2 normalised read numbers) in different tissue types of known biosynthesis genes for production of DMSP, glycine betaine and proline, with colours corresponding to panel A. **C.** Abundance of zwitterionic osmolytes measured in leaf tissue of *S. anglica.* Corresponding chemical structures are shown. Data in panels A and C represent mean ± standard deviation from six independent biological replicates.

Alt text: Diagrams and graphs labelled A to C. A and C are graphs showing abundance of different osmolytes in *Spartina anglica* leaf tissue. B is a heatmap showing expression of *Spartina anglica* osmolyte biosynthesis genes in different tissues.

Figure 6: A retrotransposon causes high-level expression of *DOX* and contributes to *Spartina anglica* making high concentrations of DMSP

A. Heatmap showing expression (DESeq2 normalised read numbers) of each *DOX* locus in different tissue types of *Spartina anglica* (Sa) and *Spartina alterniflora* (Salt). **B.** Schematic depicting each *DOX* loci of *S. anglica* and *S. alterniflora*. Blue and orange boxes indicate gene sequences (including introns and exons) and long terminal repeat (LTR) retrotransposons, respectively. Red boxes indicate the 55 bp region found uniquely in LTR^{DOX} and not in any other genes containing a similar LTR. **C.** Phylogenetic tree constructed using LTR sequences from *SaDOX6*, *SaltDOX3* and the BLAST hits from the *S. anglica* genome where the LTR was detected within 1 kb of a transcriptional start site. LTR^{DOX} of *SaDOX6* and *SaltDOX3* cluster most closely together. **D.** Multiple sequence alignment of LTR sequences from panel C showing the unique 55 bp sequence of LTR^{DOX} (shaded orange and only found in *SaDOX6* and *SaltDOX3*). Conserved bases are indicated in red. **E.** Quantification of promoter activity in *Nicotiana benthamiana* leaf tissue expressing either the 35S minimal promoter alone (p35S-min) or the 55 bp sequence of

LTR^{DOX} cloned upstream of the 35S minimal promoter (LTR^{DOX}-p35S-min). Data represent mean ± standard deviation from three independent biological replicates.

Alt text: Diagrams and graphs labelled A to E. A is a heatmap showing expression of *DOX* in *Spartina anglica* and *S. alterniflora*. B is a diagram showing *DOX* gene structures. C is a phylogenetic tree showing retrotransposon sequences from *S. anglica*. D is a multiple sequence alignment of retrotransposon sequences from *S. anglica*. E is a graph showing quantification of promoter activity in in *Nicotiana benthamiana* leaf tissue.

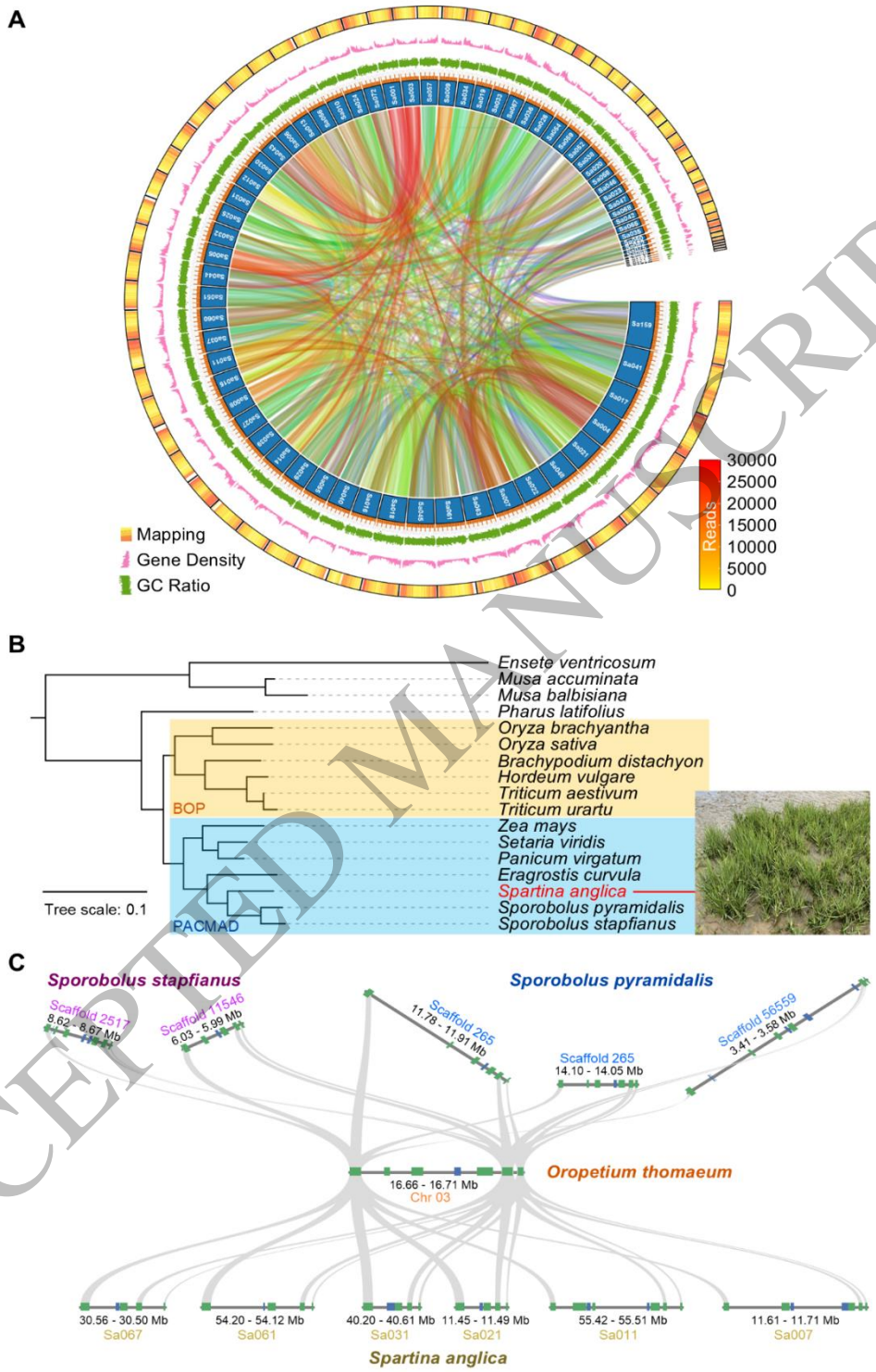

Figure 1
140x230 mm (x DPI)

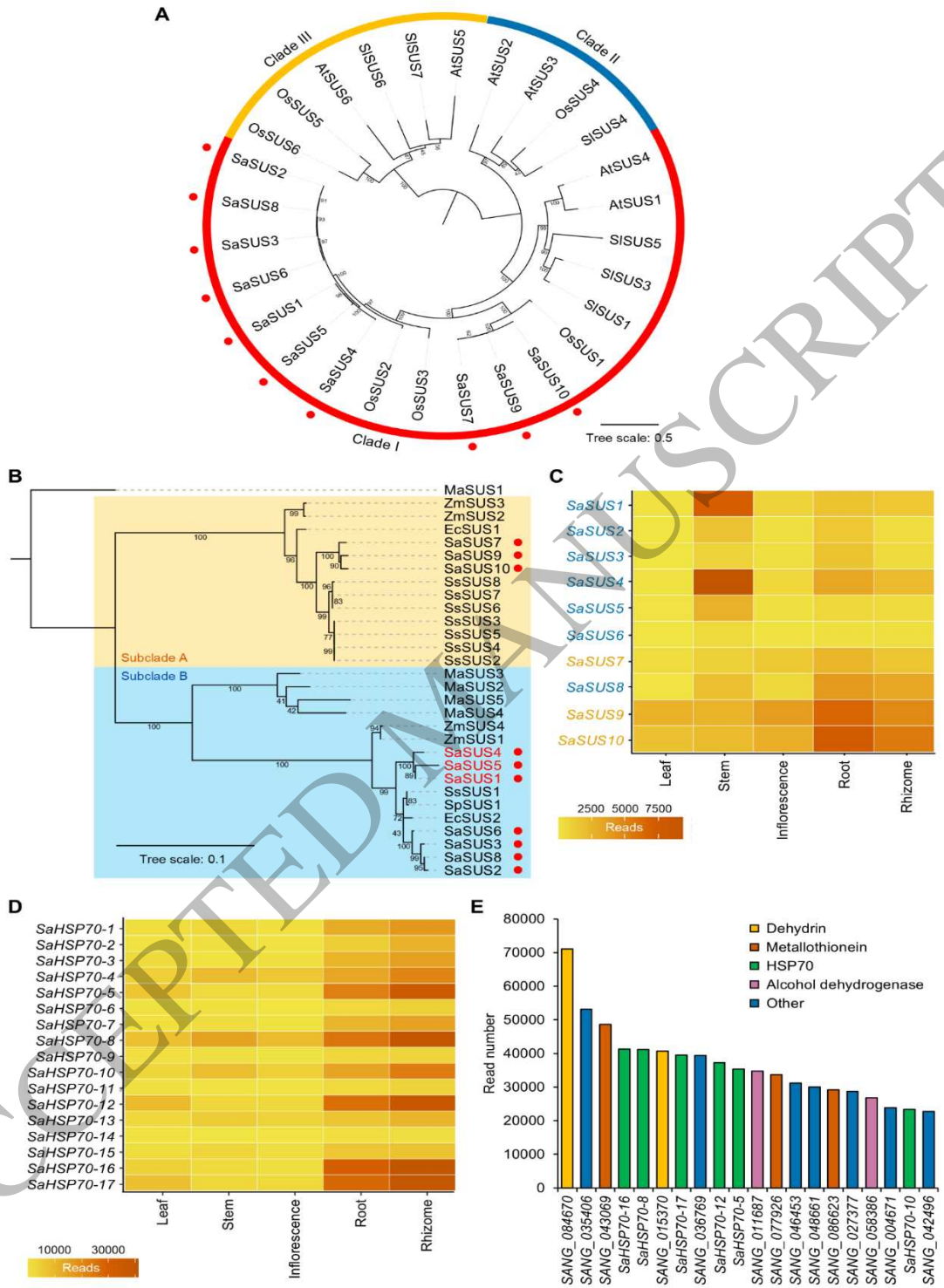

Figure 2
184x280 mm (x DPI)

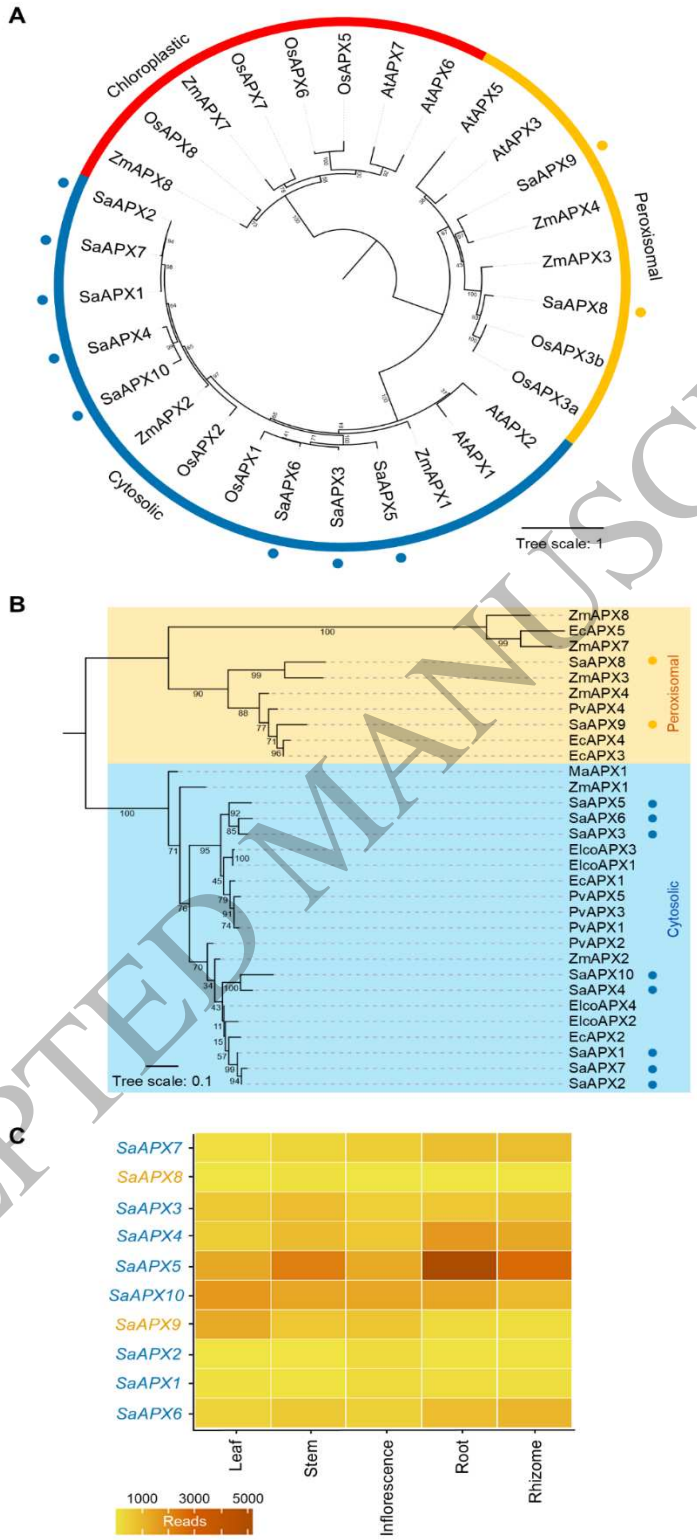

Figure 3
120x273 mm (x DPI)

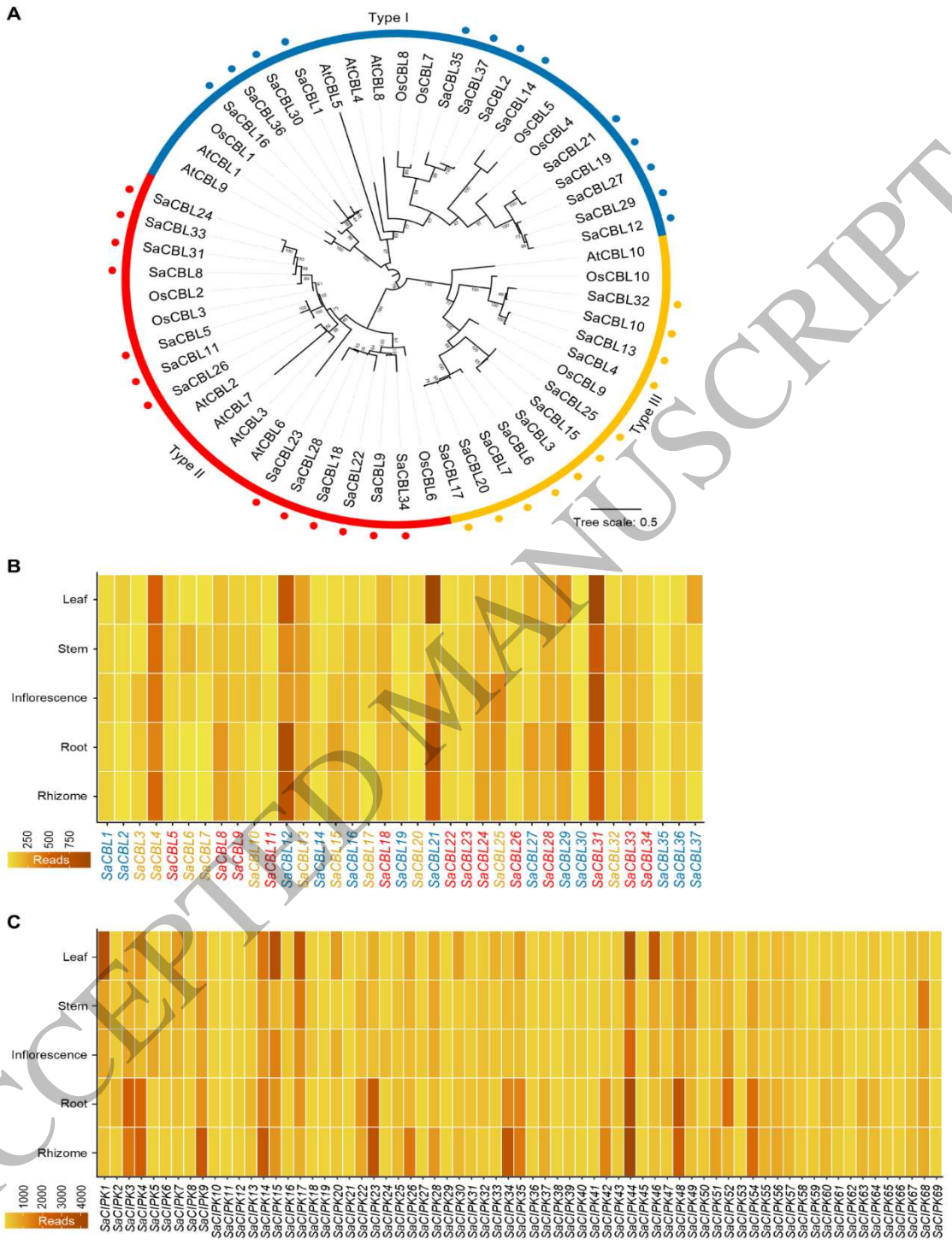

Figure 4
190x270 mm (x DPI)

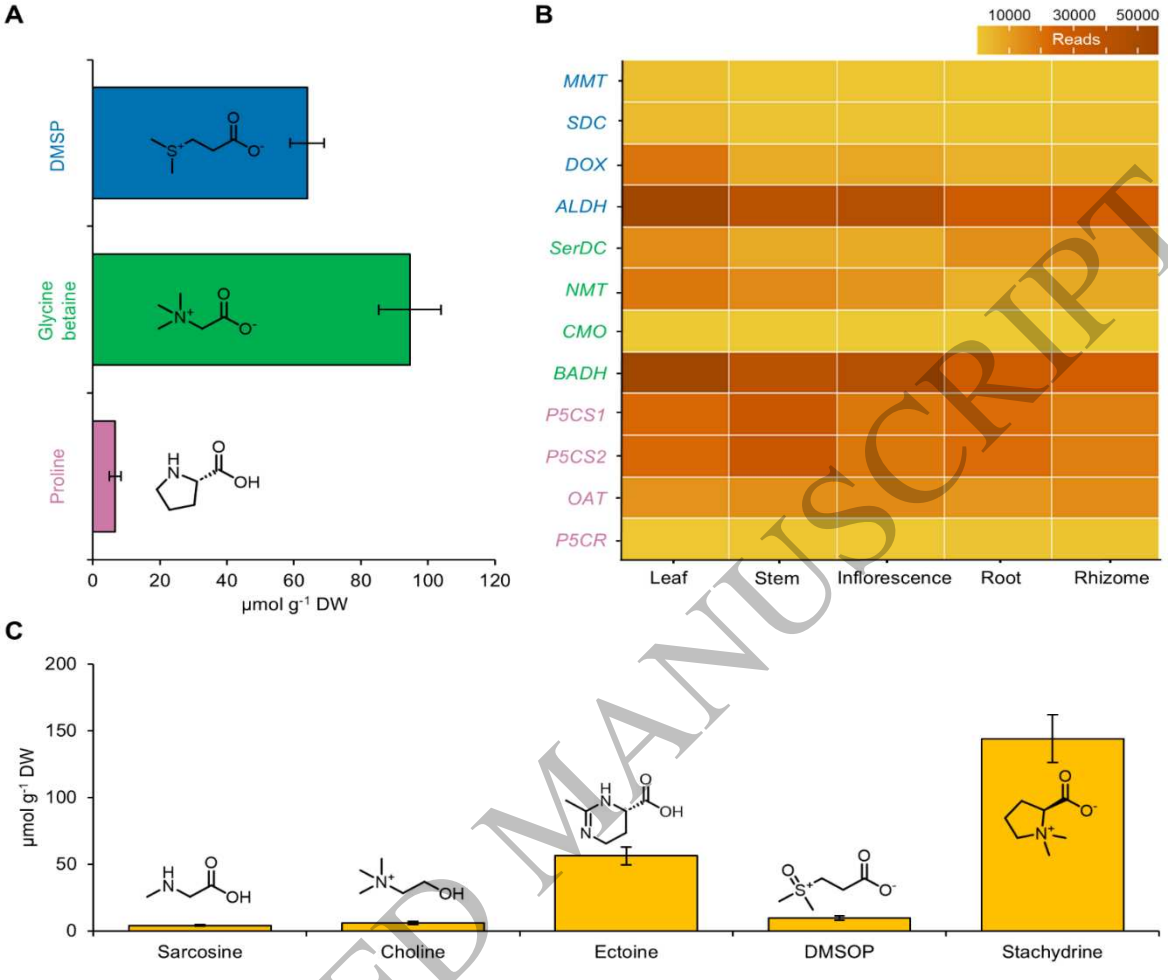

Figure 5
180x158 mm (x DPI)

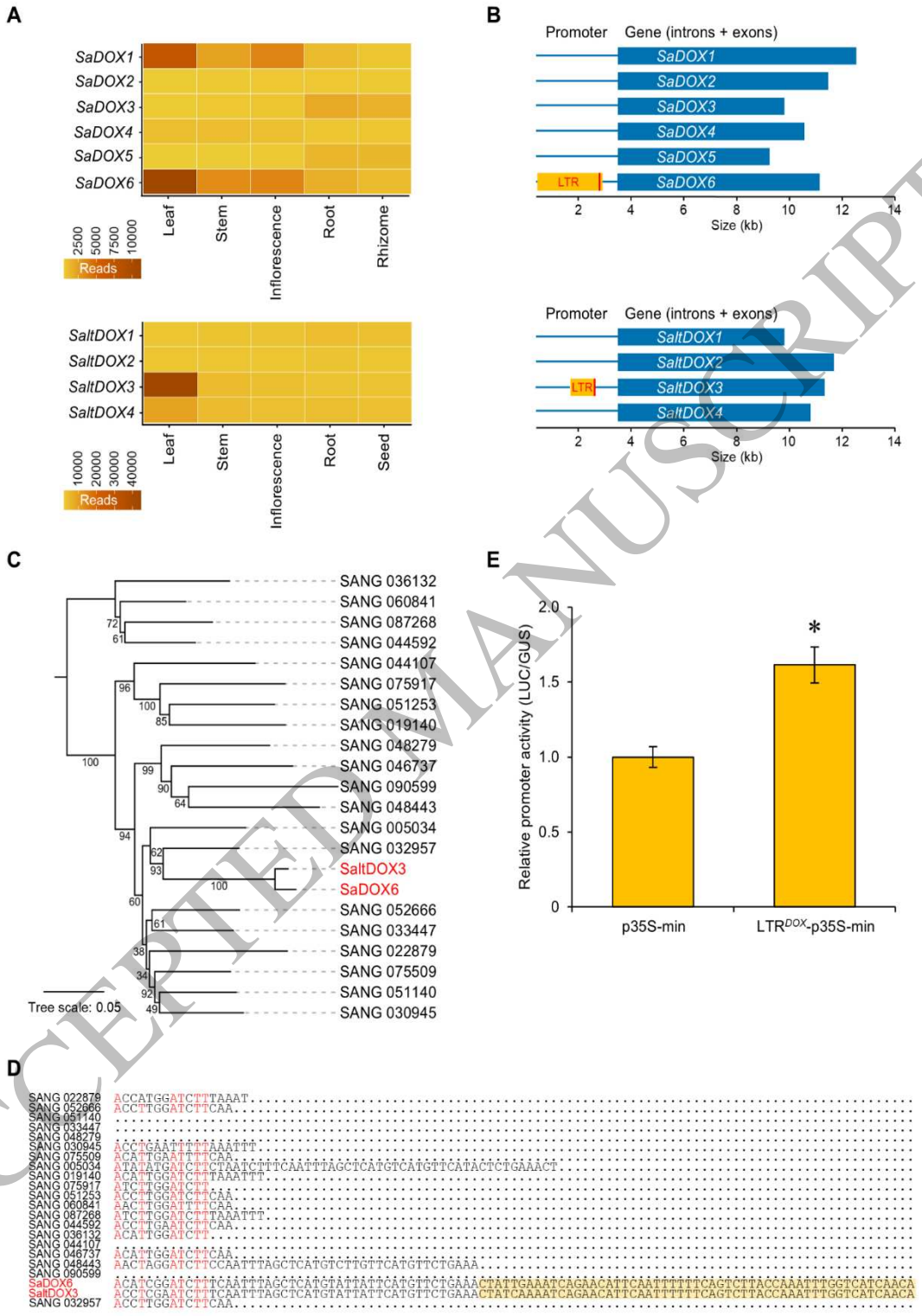

Figure 6
180x238 mm (x DPI)