Journal of Heredity, 2026, **117**, 357–365
https://doi.org/10.1093/jhered/esaf062
Advance access publication 23 September 2025
Genome Resources

American Genetic Association
OXFORD

Genome Resources

The annotated, chromosome-scale *Salicornia depressa* (American pickleweed) genome

Cooper Kimball-Rhines[1], Alice E. Kaisla[1], Scarlet Taveras Guzman[1], Emily Mello[1], Karol Vanessa Rojas Ramirez[1], Alex Harkess[2], Brook Moyers[1,*]

[1]Biology Department, University of Massachusetts Boston, Boston, MA 02125, United States, [2]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, United States

*Corresponding author: Biology Department, University of Massachusetts Boston, Boston, MA 02125, United States. Email: brook.moyers@umb.edu

Abstract

Salicornia depressa (American pickleweed) is the most widespread member of its salt-loving genus in North America. Pickleweeds typically colonize high marsh areas that tides make highly saline. Most other salt marsh plants cannot withstand the same salt pressure, so these bare patches are free from competitive pressures. Understanding the genetic origin of American Pickleweed's high salinity adaptations has potential application to salt-tolerant agriculture and salt marsh conservation but requires genome resources to study. *S. depressa* is a tetraploid species, which presents unique challenges to traditional genome assembly pipelines. We present a high-quality, chromosome-scale reference genome of *S. depressa* and the pipeline we used to assemble it. Our reference is phased to each of the tetraploid's ancestral subgenomes with a scaffold N50 of 69.3 Mb, BUSCO completeness of 96.4%, and k-mer completeness of 98.4%. The subassemblies are evenly split, with subassembly A containing 56.7% and subassembly B containing 59.1% of genomic k-mers. Our gene annotation identifies 80,883 genes and our methylome annotation contains 2.5 million methylated cytosines. We find that gene and methylation density are negatively correlated across the genome. We also assembled and annotated a chloroplast assembly which includes all expected photosystem, tRNA, and rRNA genes. We provide a guide to our successful assembly pipeline involving >30 programs. Our reference and annotation join resources for three other *Salicornia* species, allowing global scale ecological-evolutionary studies.

Key words: genome assembly, methylome, polyploidy, salt marsh, transcriptome

Introduction

Coastal ecosystems support a disproportionate amount of human life worldwide with salt marshes provisioning the most services of any ecosystem by area (Adams et al. 2021). Human coastal communities benefit from flood mitigation, erosion protection, pollutant filtration, and recreation; animals use marshes for shelter, food, and to raise their young. While dependent on marshes for these services, coastal cities are the leading causes of marsh destruction through development and urbanization (Bertness et al. 2002). Intact urban marshes face new challenges to their survival, particularly eutrophication and sea level rise (Krause et al. 2020; Cahoon et al. 2021).

Marsh plants directly influence the structure and stability of their ecosystems. Roots physically hold the marsh together; their slow decay results in high density carbon sequestration (Drake et al. 2015). Salt marshes are set in a constantly shifting pattern of area loss and reclamation—a cycle that is dependent on a small set of plant species. Stronger tides at the full and new moon wash marine wrack into the high marsh, where it remains to smother plants that were growing underneath. This creates a patchwork of bare areas in the high marsh that build up salinity through evaporation and salt spray. These areas exhibit salinities greater than 100 parts per thousand, more than three times that of salt water and higher than the tolerance of many salt marsh plants (Bedre et al. 2016).

To survive their harsh environment, plants in the genus *Salicornia* (commonly pickleweeds or samphire) utilize membrane proteins that take-up and pump sodium ions into the plant's vacuoles (Salazar et al. 2024). This strategy safely sequesters sodium into vacuoles, adjusting cellular osmotic pressure. More than several generations, sequestration of sodium into tissue is enough to lower salinity of bare patches so later successional species, including cord grasses (*Spartina*) and rushes (*Juncus*), may grow (Chapman 1940; Farzi et al. 2017). Reportedly by this same mechanism, *Salicornia* species tolerate and hyperaccumulate a wide range of metals, including lead, copper, nickel, zinc, cadmium, arsenic, vanadium, and selenium (Sharma et al. 2010). Salt marshes are often highly polluted with heavy metals, particularly those near urban centers (Zhang et al. 2020). *Salicornia* species may act as bioremediators of these metals, presenting an alternate solution for long-term marsh cleanup that does not involve destruction of the ecosystem (Milic et al. 2012; Mukherjee et al. 2021).

The *Salicornia* genus (Salicornioideae, Amaranthaceae) currently includes 25 to 30 species with a largely temperate

Received on 21 May 2025; accepted on 30 August 2025

global distribution. Attempts to taxonomically classify *Salicornia* species have encountered difficulty reconciling historic observations, herbarium specimens, and modern genetic techniques. The current taxonomic organization groups species based on internal transcribed spacer (ITS) sequences (Kadereit et al. 2007). Numerous polyploidization and reduction events appeared in the diversification of the genus; extant diploids, tetraploids, and a decaploid species remain, sometimes inhabiting the same areas. Their mix of cryptic polyploidy and taxonomic mystery, along with substantial phenotypic plasticity, make *Salicornia* an intriguing model for evolutionary studies.

Ecologically, salt marshes are a classic model for theories of competition and stress gradients, in part because there is relatively low species diversity in these ecosystems worldwide (Emery et al. 2001). However, there has been little use of salt marshes to study evolutionary biology. Genetic resources for salt marsh species have trailed behind those available for other ecosystems. We are beginning to close that gap now, with reference genomes for three *Salicornia* species (*Salicornia ramosissima, Salicornia europaea*, and *Salicornia bigelovii*), two *Sueda* species (*Sueda aralocaspica* and *Sueda glauca*), and one *Spartina* species (*Spartina alterniflora*) released more than the last 5 years. The resources for ecosystem-scale genomic studies in salt marshes are becoming more available, presenting an opportunity to link fundamental ecological studies with evolution. These plant systems have applications to national and global priorities, including carbon sequestration, addressing heavy metal pollution, and mitigating ecosystem destruction.

Here, we present the first de novo reference genome of *S. depressa*, an allotetraploid and the most widespread member of the genus in North America (Fig. 1a). Our assembly is based on Pacific Biosciences (PacBio) HiFi and Oxford Nanopore long reads and scaffolded with Phase Genomics Proximo Hi-C data. Alongside it, we provide a chloroplast assembly, transcriptome, and methylome annotation. The inherent complexity in many non-model plant organisms makes the assembly process challenging, with many traditional tools unusable for such species (Li and Harkess 2018). For polyploids, assemblers often have trouble differentiating parental haplotypes and duplications, necessitating careful and costly tuning to match the individual species. To assist in future projects, we present a thorough guide to our pipeline with notes on program installation, parameter tuning, and all code files we used to create our assemblies and annotations.

Methods

Biological materials

We sampled shoot tip tissue from large, wild grown *S. depressa* plants in Savin Hill Cove, Boston, Massachusetts twice. In 2021, we collected material to profile the genome and annotate the transcriptome; in 2022, we located and sampled three reference candidates. One plant was selected as the reference individual after assessing the quality of extractions, whereas the other two provided additional data for transcriptome annotation. The reference plant's location is GPS tagged as 42.31162, −71.04412. We also grew plants from seed collected in 2022 from the same population in a nitrogen stress experiment and collected root, shoot, and reproductive tissue for RNA sequencing. Immediately after harvesting, we flash-froze the tissue in liquid nitrogen and stored it at −80°C until extraction.

Nucleic acid library preparation

To produce whole genome shotgun libraries for genome profiling, we sampled shoot tip tissue from three 2021 plants and extracted DNA with a Qiagen Plant Pro Kit (Cat. #69204). The UMass Boston Center for Personalized Cancer Therapy created a whole genome shotgun library using the KAPA Hyper Prep Kit performed on a Beckman Biomek FXp Laboratory Automation Workstation. To extract high molecular weight DNA from the reference candidate plants, we followed the Takara Bio NucleoBond HMW DNA Kit (Cat. #740160.20). In brief, we homogenized tip tissue in liquid nitrogen, bound the DNA to a column, then eluted and precipitated the DNA in isopropanol. This extraction method produced a high-quality extract with a Qubit-quantified concentration of 171 ng/μl and Femto Pulse-quantified average fragment size of 50 kb for our reference plant. Hudson-Alpha produced a PacBio HiFi library using DNA that was sheared using a Diagenode Megaruptor 3 instrument. Libraries were constructed using SMRTbell Template Prep Kit 2.0 and tightly sized on a SAGE ELF instrument (1 to 18 kb) to a final library average insert size of roughly 20 kb. We also used the Oxford Nanopore Native Barcoding Kit 24 V14 (SQK-NBD114.24) to prepare a library from the same DNA for Nanopore long read sequencing. To aid in scaffolding, we prepared a chromatin crosslinked extract from the same reference tissue with the Phase Genomics Proximo Hi-C plant kit (KT3040). For transcriptome annotation, we extracted RNA from shoot tissue of the reference individual and shoot, root, and reproductive tissue from 35 plants collected wild or grown under nitrogen pollutant stress. All plants were from the same Savin Hill Cove population as the reference. We extracted bulk RNA using a Qiagen RNeasy Plant Kit (Cat. #74904) and treated extracts with DNase and the RNeasy MinElute Cleanup Kit (Cat. #74204) (see JGI DNase Treatment 20 120 909). The Joint Genome Institute produced two Isoseq long read libraries for the reference plant extract using the SMRTbell Barcoded Adapter Plate 3.0 and the SMRTbell Express Template Prep Kit 2.0 (PacBio). JGI size selected one of the Isoseq libraries at a lower cutoff fragment size of 1.5 kb. RNA from all other samples were made into Illumina Regular (RNA) libraries (SOP 1070.4).

DNA sequencing

The UMass Boston Center for Personalized Cancer Therapy sequenced our three whole genome shotgun libraries on an Illumina NextSeq 2000 with P2 chemistry. The runs produced a total of 500 million 100 bp single end reads for use in genome profiling. Hudson-Alpha sequenced the PacBio library on two SMRT cells, which produced 535 and 514 Gb of raw sequencing data. CCS processing left 31.4 and 31.7 Gb of HiFi data for use in the assembly. We sequenced the Nanopore library on a single MinION Mc1k R10.4.1 flow cell, which yielded 4.74 Gb of sequencing data. The UMass Boston Center for Personalized Cancer Therapy sequenced our Hi-C library on an Illumina Nextseq 2000 with P2 chemistry. We obtained 300 million paired end reads at 100 bp each. JGI sequenced both long read RNA libraries on a Pacific Biosystems' Sequel IIe sequencer using SMRT Link 12.0, tbd-sample dependent sequencing primer, 8M v1 SMRT cells, and Version 2.0 sequencing chemistry with 1 × 1800 sequencing movie run times. The short read libraries

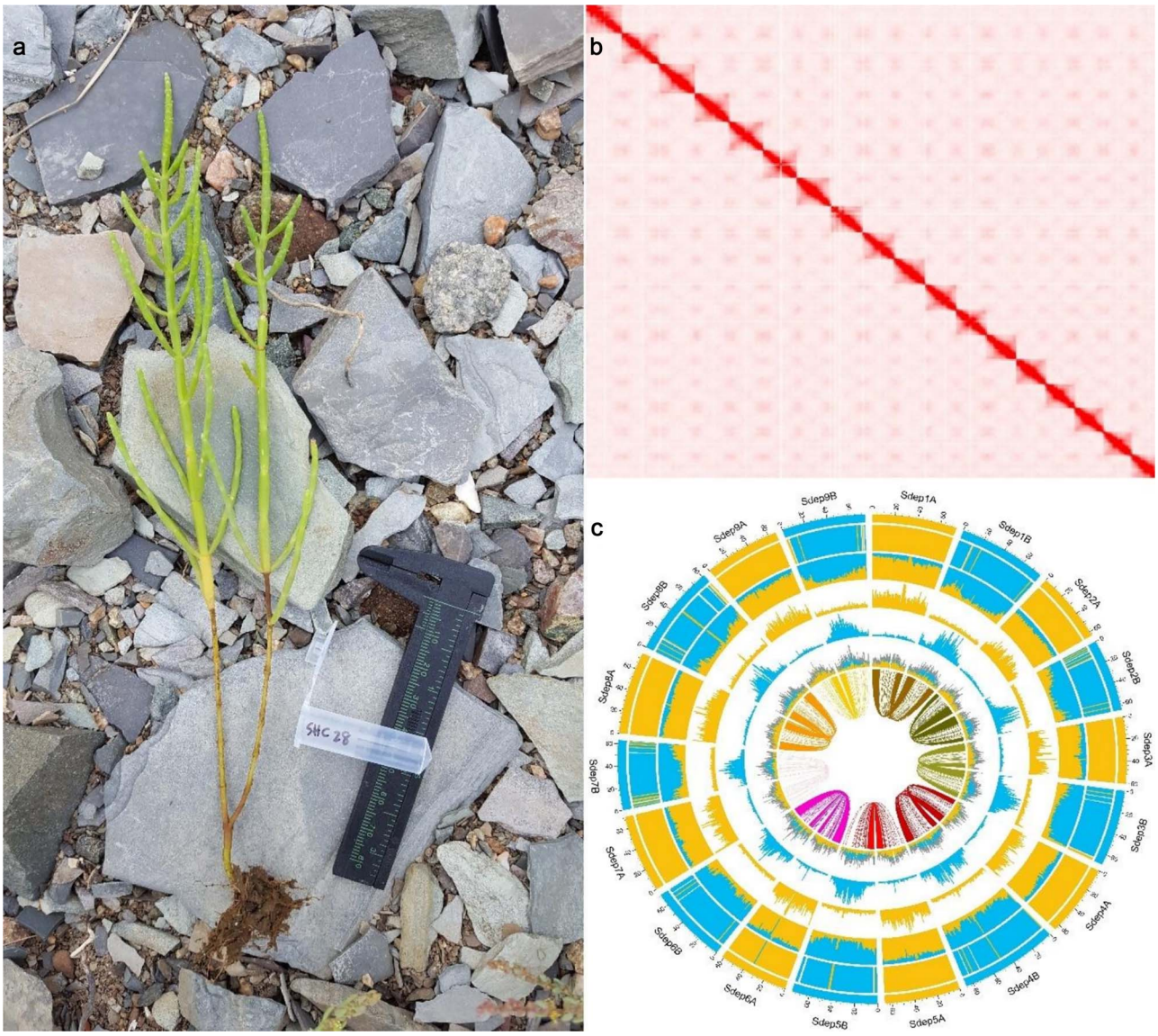

Fig.1. *Salicornia depressa* chromosome-scale scaffold assembly. (A) A representative *S. depressa* adult with measuring instrument. (B) Hi-C contact map of chromosome-scale nuclear genome scaffolds. Boxes in the map denote continuous genome regions corresponding to chromosomes. (C) SubPhaser Circos plot of k-mer phased subgenomes. Rings from the outside-in correspond to (1) karyotypes, (2) subgenome assignment, (3) normalized proportion of each subgenome, (4) density of subgenome A-specific k-mers, (5) density of subgenome B-specific k-mers, (6) density of LTR-RTs, and (7) homologous k-mer blocks.

were sequenced 2 × 151 using the Illumina NovaSeqX platform.

Genome assembly

We first assessed the quality of our reads using *fastqc* (Andrews 2010) and, where appropriate, trimmed adaptors with *Cutadapt* (Martin 2011). We visualized the quality and length of our long reads using *Nanoplot* (De Coster and Rademakers 2023). To profile the genome, we processed our whole genome shotgun library into k-mers using *KMC* (Kokot et al. 2017), which produced input files for *GenomeScope 2.0* and *smudgeplot* (Ranallo-Benavidez et al. 2020). For contig assembly, we provided *hifiasm* with PacBio HiFi reads, Nanopore long reads, and Hi-C paired end reads (Cheng et al. 2021; Cheng et al. 2022). More than several assemblies, we tuned the similarity threshold parameter (-s), which controls duplicate purging, to 0.35. Due to our genome's allotetraploidy, we manually set the k-mer peak of homozygous coverage to 93× (–hom-cov 93) and determined that the program was best run using the polyploid setting (–n-hap 4). Even with parameter tuning, the four phased haplotypes were skewed in their sizes, so we moved forward with the unphased, unpurged primary unitigs. We used *blobplot* to assess contamination in our unitigs (Challis et al. 2020). We used *HapHiC* to scaffold our genome, providing both an alignment of Hi-C paired end reads to our contigs and the assembly gfa file with a target chromosome number parameter set to 18 (Zeng et al. 2023). We visualized our scaffolds using *Juicertools* and *juicebox-web* to create a contact map (Durand et al. 2016a; Durand et al. 2016b). We used *SubPhaser* to group our scaffolds into the two allotetraploid ancestral lineages (Jia et al. 2022). We polished our final assembly with *racon* (Vaser et al. 2017) and performed masking with *repeatmasker* (Smit et al. 2013-2015). We used *BUSCO* and *Merqury* to assess assembly quality (Simão et al. 2015; Rhie et al. 2020). To order

and name our chromosomes, we mapped BUSCO orthologues between our assembly and the Quinoa reference genome which was chosen because it is a well-studied relative within the *Chenopodiaceae* family.

Genome annotation

We performed transcriptome annotation using the long read version of *BRAKER3* to produce gene hints with protein, short read RNA, and long read RNA evidence. We produced gene sets with *AUGUSTUS* and *GeneMark* before combining evidence streams using *TSEBRA* (Lomsadze et al. 2005; Stanke et al. 2006; Gotoh 2008; Stanke et al. 2008; Li et al. 2009; Barnett et al. 2011; Iwata and Gotoh 2012; Lomsadze et al. 2014; Buchfink et al. 2015; Hoff et al. 2016; Hoff et al. 2019; Bruna et al. 2020; Bruna et al. 2021; Gabriel et al. 2021). We used *BUSCO* to create a new set of training parameters for AUGUSTUS based on initial default settings. We performed an ontology annotation using *Mercator* and annotated protein domains using *InterProScan*. We also produced a methylome annotation from our Nanopore data using *Dorado* and *Modkit* (Oxford Nanopore Technologies 2023; Oxford Nanopore Technologies 2024). We visualized the completed assembly and annotations with the R package *circlize* (Gu et al. 2014). Finally, we assembled a chloroplast reference using *Trycycler* by creating a consensus of several assemblies from *flye* and *unicycler* (Wick et al. 2017; Kolmogorov et al. 2019). We annotated the chloroplast using *CPGAVAS2* (Shi et al. 2019). See Table 1 for version numbers of all programs. For all custom code files, see https://github.com/coopermkr/sdepressaAssembly.

Results

Members of the *Salicornia* genus have high variation in their genome size, in part because of multiple polyploidization and reduction events. *S. bigelovii*, the sister tetraploid species to *S. depressa*, has a genome size of 2.0 Gb. The next most closely related member of the genus, the diploid *S. europaea*, has a 515 Mb reference genome. Profiling results from Smudgeplot using whole genome shotgun data indicate that *S. depressa* is a diploid—a result that contravenes previous chromosome squash results from populations in the same region (see Supplementary materials; Wolff and Jefferies 1987). GenomeScope 2.0 requires ploidy as a provided parameter for genome size and heterozygosity estimation. If we indicate that *S. depressa* is a diploid, GenomeScope 2.0 estimates the haploid genome size at 565 Mb, in line with the diploid *S. europaea*. However, the estimate of heterozygosity is nearly 7%, which would make *S. depressa* the most highly heterozygous diploid plant recorded to date—a mean feat for a reported obligate-selfing species. Profiling the species as a tetraploid provides more reasonable results, with 5.75% of heterozygosity attributed to differences between allotetraploid ancestral genomes (aabb) and a haploid genome length of 282 Mb (see supplementary materials).

Proper assembly of the tetraploid haplotypes requires long read data, which we gathered using both PacBio and Oxford Nanopore sequencing platforms. Dotplot visualization of our PacBio and Nanopore data shows a consistent pattern between the two platforms: PacBio reads are shorter but far higher quality than Nanopore reads (see supplementary materials). Incorporating reads from both platforms into an assembly drastically improved contiguity while maintaining sequence quality. Our initial unphased primary unitig assembly, based on PacBio, Nanopore, and Hi-C paired end short read data, totaled 1.39 Gb across 4,177 contigs, with a contig N50 value of 1 Mb (Table 2). Before scaffolding, we screened our contigs for contaminants with Blobplot using alignments against the Uniprot and BUSCO databases. Results showed matches between our contigs and the clades Streptophyta and Artverviricota (see supplementary materials). The former includes all land plants, which our species belongs to, while the Artverviricota include families of common plant endogenous retroviruses. We concluded that our contigs were free from substantial contamination and suitable for use in scaffolding.

During the contig assembly process, Hifiasm uses Hi-C data to bias an assembly toward a successful scaffold. We included Hi-C data in both the contig assembly and scaffold assembly stages to make full use of this feature. HapHiC requires a target number of scaffolds based on the number of chromosomes of the species. We supplied it with $N = 18$ to capture the allotetraploid lineages as a consensus sequence of our reference individual's haplotypes. The result produced 18 chromosome-scale scaffolds and 1,346 unplaced contigs with a scaffold N50 of 69.3 Mb, L50 of 10, and total genome length of 1.38 Gb (Table 2). A contact map of the assembly confirms that there are no misjoins or chimeric chromosomes (Fig. 1b). We next aligned our assembly against the chromosome-scale *S. ramosissima* reference to determine synteny. We assigned full scaffolds that aligned to the same *S. ramosissima* scaffold as chromosome pairs, which we supplied to SubPhaser to determine the groupings of our tetraploid lineages with k-mers (Fig. 1c). To ensure collapsed regions of our assembly were filtered out, we assessed coverage of our HiFi data across the 18 chromosome-scale scaffolds. On average, our scaffolds had 48× coverage by PacBio reads with little variation across scaffolds, indicating absence of collapsed assemblies. K-mer scores, as calculated by Meryl and Merqury, indicate 98.4% completeness across the assembly (see Supplementary materials), while BUSCO scores against the Eudicot V5.7.1 dataset show 96.4% completeness, with 14.1% single-copy, and 82.3% double-copy (Table 2). Finally, our scaffolds show high synteny with the most recently published tetraploid quinoa (*Chenopodium quinoa*) assembly, a somewhat closely related member of the Amaranthaceae family (Rey et al. 2023). We mapped BUSCO orthologues against quinoa to order and name our chromosome-length scaffolds (Fig. 1d).

To annotate our complete assembly, we generated gene hints from protein alignments, short read RNA, and long read RNA. The *BUSCO*-trained *BRAKER* and *AUGUSTUS* pipeline identified 80,883 genes across the genome, including both tetraploid subgenomes (see Supplementary materials for *AUGUSTUS* parameters). Gene ontology analysis with Mercator annotated 77,216 of these transcripts and classified 22,920. Transcripts occupied 87% of Mercator's bins, including 16% of clade-specific metabolism bins (see Supplementary materials). Of these transcripts, 66,181 fall on the 18 chromosome-scale scaffolds. Modkit identified 483,332 5hmC methylated loci and 1,987,742 5mC methylated loci across the genome. The number of genes and methylated cytosines in 1 Mb windows are negatively correlated (Pearson's $r = -0.226$, $P = 2e\text{-}16$). We visualized the completed genome with its annotations and identified orthologues between subgenomes using the R package *circlize* (Fig. 2a).

Our genome Hi-C contact map shows that many of the unplaced unitigs in the assembly align closely with one

Table 1. Assembly and annotation pipeline programs.

Assembly stage	Software and options	Version	Reference
Package management	miniconda	22.11.1	Anaconda Software Distribution (2020)
Read quality control	fastqc	0.11.9	Andrews (2010)
Read visualization	NanoPlot	1.32.1	De Coster et al. (2023)
Adaptor trimming	cutadapt	4.6	Martin (2011)
K-mer counting	kmc (-k21 -ci1 -cs10000)	3.1.1	Kokot et al. (2017)
Genome profiling	smudgeplot	0.2.3	Ranallo-Benavidez et al. (2020)
Ploidy profiling	GenomeScope	2.0	Ranallo-Benavidez et al. (2020)
De novo assembly	hifiasm (–hom-cov 93 -s 0.35 –s-base 0.35)	0.19.8	Cheng et al. (2022)
Contaminant detection	BlobToolKit	4.3.7	Challis et al. (2020)
	DIAMOND (blastx –sensitive)	2.1.7	Buchfink et al. (2015)
Scaffolding	minimap2 (-ax sr)	2.26	Li (2021)
	samtools	1.14	Danecek et al. (2021)
	HapHiC	1.0.2	Zeng et al. (2023)
Assembly visualization	JuicerTools	1.2	Dudchenko et al. (2018)
	juicebox-web	2.3.5	Durand (2016a)
Phasing	SubPhaser	1.2.6	Jia et al. (2022)
Polishing	minimap2 (-ax map-hifi)	2.26	Li (2021)
	racon	1.4.3	Vaser et al. (2017)
Masking	repeatmasker	4.1.2	Smit et al. (2013-2015)
QAQC	BUSCO	5.4.4	Simão et al. (2015)
	Meryl	1.4.1	Rhie et al. (2020)
	Merqury	1.3	Rhie et al. (2020)
Synteny	pafr	0.0.2	Winter (2025)
	circlize	0.4.16	Gu et al. (2014)
Chloroplast assembly	minimap2 (-ax map-hifi)	2.26	Li (2021)
	samtools	1.14	Danecek et al. (2021)
	bedtools	2	Quinlan and Hall (2010)
	BBMap	38.90	Bushnell (2014)
	flye	2.9	Kolmogorov et al. (2019)
	unicycler	0.5.0	Wick et al. (2017)
	racon	1.4.3	Vaser et al. (2017)
	blast-plus	2.13.0	Camacho et al. (2009)
	Trycycler	0.5.5	Wick et al. (2021)
Chloroplast annotation	CPGAVAS2		Shi et al. (2019)
Methylome annotation	Dorado	0.5.1	ONT (2023)
	Modkit	0.3.1	ONT (2024)
Transcriptome annotation	BRAKER and AUGUSTUS (see Supplementary materials for custom parameters)	3.0.8, 3.5.0	Barnett et al. (2011); Bruna et al. (2020, 2021); Buchfink et al. (2015); Gabriel et al. (2021); Gotoh (2008); Hoff (2016, 2019); Iwata and Gotoh (2012); Li et al. (2009); Lomsadze et al. (2005, 2005); Stanke et al. (2006, 2008)

Table 2. *Salicornia depressa* assembly statistics.

HiFi read coverage[a]	44x					
Number of contigs	4177					
Contig N50 (Mb)	1					
Longest contig (bp)	12,979,358					
Number of scaffolds	2831					
Scaffold N50 (Mb)	69.3					
Scaffold L50	10					
Longest scaffold (bp)	86,372,990					
Size of final assembly (Gb)	1.38					
Gaps per Gbp	662.3					
k-mer completeness	98.40%					
BUSCO completeness[b]	C	S	D	F	M	
	96.4	14.1	82.3	0.3	3.3	n: 2326
Chloroplast length (bp)	153,298					

[a]Reported coverage is the average coverage of PacBio HiFi reads aligned against the assembled chromosome-scale scaffolds. [b]BUSCO scores: (C)omplete, (S)ingle, (D)uplicated, (F)ragmented, and (M)issing genes. *n* is total number of genes in the dataset.

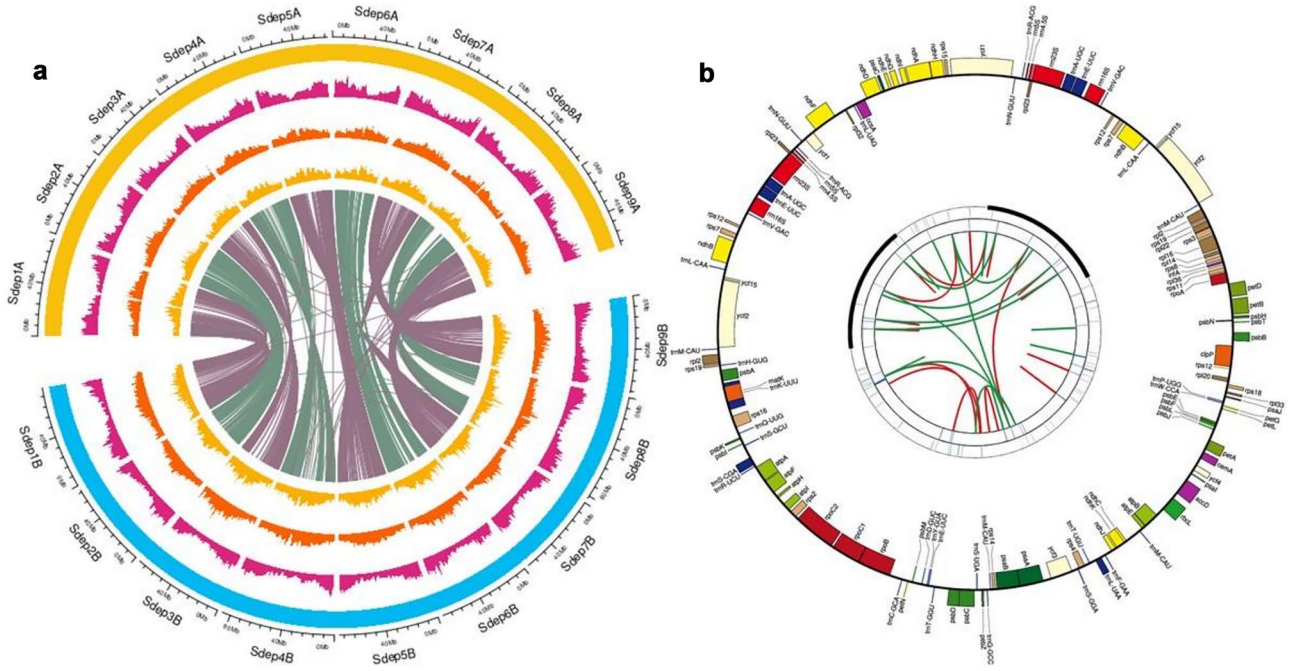

Fig. 2. Nuclear genome and chloroplast annotations. (A) Circos plot of the 18 chromosome-scale scaffolds assigned to subgenome and named based on gene synteny with quinoa reference genome. Rings from the outside-in denote (1) chromosome name, (2) subgenome assignment, (3) gene density, (4) 5mC methylation density, (5) 5hmC methylation density, and (6) BUSCO position ribbons. (B) Chloroplast assembly annotated by CPGAVAS2. Rings from the outside-in denote (1) open reading frames of identified genes, (2) long tandem repeats, and (3) microsatellites.

another. Further investigation indicated that these unitigs were pieces of the chloroplast. Rather than attempt to join these pieces together, we curated ten starting read datasets by subsampling an alignment of our PacBio data against the published *S. bigelovii* chloroplast. We used *Flye* and *Unicycler* to create an assembly from each read set and generated a consensus chloroplast sequence with *Trycycler*. Our assembly shows a circularized single contig of 153,300 bp with two inverted repeats, and one long and one short single-copy region. We annotated the chloroplast using CPGAVAS2 (Shi et al. 2019), which included all major tRNA, rRNA, and photosystem genes (Fig. 2b).

Discussion

Genome assembly methodologies for complex plants lack standardization. In part, the speed of development in sequencing and computational technologies makes the goal of standardization for plant genome assembly impossible. Detailed documentation of an assembly pipeline can both capture the current state of methods in the field and make the process more reproducible. In pursuit of these goals and to assist others interested in pursuing non-model organisms with complex genomes, we present a GitHub repository containing our code files and a wiki walkthrough of our program and parameter choices (https://github.com/coopermkr/sdepressaAssembly).

Using our custom pipeline, we created a high-coverage, highly contiguous reference genome with transcriptome and methylome annotations for the species *S. depressa* (American pickleweed). With this report, reference genomes of four *Salicornia* species are now publicly available, all of which contain chromosome-scale scaffolds. These species include two diploids with largely European and Mediterranean ranges and two North American tetraploids. The scaffold N50 values of the assemblies range from 51 Mb in the diploid *S. europaea* to 98 Mb in the tetraploid *S. bigelovii* (Salazar et al. 2024). GC content is similar across the four assemblies, with the lowest at 34.77% in diploid *S. ramosissima* and the highest at 36.98% in *S. bigelovii* (Mian et al. 2024). All the assemblies are highly complete, with *S. bigelovii* at 95.3% overall BUSCO completeness and the other three grouped within a tenth of a percentage point around 96.4%. As is expected, single-copy orthologues make up almost all of those found in the diploid species, while the tetraploids primarily have duplicates. Genome size varies widely across the four references, with assembled haploid (1 N) genome lengths of 516 Mb in *S. europaea*, 529 Mb in *S. ramosissima*, 1,380 Mb in *S. depressa*, and 2,026 Mb in *S. bigelovii*. The difference in genome size between the two tetraploids indicates an evolutionary gulf possibly driven by repetitive elements that may help date the divergence of the species in North America.

The *Salicornia* genus makes a compelling model for the study of polyploidization and genome reduction, as it appears both events have taken place multiple times in the genus's evolutionary history. The most recent taxonomy of the genus has shifted to group species by their ploidy but leaves some paraphyletic diploid and tetraploid assignments (Kadereit et al. 2007). A new taxonomy based on reference genomes would potentially resolve the genus and may indicate that these species are prone to polyploidy and reduction at higher rates than expected. This could provide insight into this mode of plant diversification and the evolution of multiple ploidy states.

Salt marshes have long been a system for investigating basic theories of competition and coexistence, making *Salicornia* a promising system to add an evolutionary angle to traditional ecological studies. Our reference represents the most common member of the genus in northern North America

and most widespread latitudinally, with *S. bigelovii* more common on both coasts of the southern United States and Mexico. *S. europaea* is most common in Northern Europe but also occurs across coastal Europe into western Asia. *S. ramosissima* has the smallest natural range, primarily growing on the Iberian Peninsula. Taken together, these four reference species represent a wide latitudinal range with a long history of intercontinental dispersal. Three of the four species have gene annotations, while ours is the only reference to also have a methylome annotation.

Understanding the genetic, transcriptomic, and methylation underpinnings of adaptation to salt stress has many potential applications. Previous studies of other members of the genus indicate that *Salicornia* species have a unique molecular approach to surviving saline environments. Investigating this mechanism in *S. depressa* and comparing it to other salt marsh species may contribute to salt-resistant agriculture. The species may also be useful in bioremediation for heavy metals, as it appears this mechanism is not specific to sodium. We hope that providing a complete genome with annotations will support investigations like these.

Acknowledgments

We would like to thank Kerrie Barry for coordinating all RNA sequencing provided by JGI. We would also like to thank the UNC Basic Assembly and Annotation of Genomes Workshop held in 2022 for thorough information on genome assembly processes, Rebecca Povilus for advice on chromosome squashes, and Jill Wegrzyn for advice on annotation.

Author contributions

Cooper Kimball-Rhines (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing—original draft, Writing—review & editing), Alice E. Kaisla (Conceptualization, Investigation, Resources, Writing—review & editing), Scarlet Taveras Guzman (Investigation, Resources), Emily Mello (Investigation, Resources), Karol Vanessa Rojas Ramirez (Investigation, Resources), Alex Harkess (Conceptualization, Methodology, Project administration, Software, Supervision, Writing—review & editing), Brook Moyers (Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing—review & editing).

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

Funding

This work was supported by a National Science Foundation Graduate Research Fellowship [Award #2235034]; the National Science Foundation REU grant DBI-1950051 to R. Skvirsky at UMass Boston; a Joseph P. Healey Research Grant to B. Moyers; and a Nantucket Biodiversity Initiative Climate Change and Island Resilience Grant; The work [proposal: 10.46936/10.25585/60008872] conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231.

Data availability

We have deposited the primary DNA and RNA sequences underlying our analyses with metadata in NCBI SRA PRJNA1248027 and PRJNA1244426. The final genome assembly is available under the accession NCBI JBNZCH000000000 and transcriptome and methylome annotation files are available at: 10.5281/zenodo.17123225.

References

Adams JB, Raw JL, Riddin T, Wasserman J, Van Niekerk L. Salt marsh restoration for the provision of multiple ecosystem services. *Diversity-Basel*. 2021;**13**:Article 680. https://doi.org/10.3390/d13120680

Anaconda Software Distribution. *Anaconda Documentation*. Anaconda Inc. 2020. Retrieved from https://docs.anaconda.com/

Andrews S. *FastQC: a quality control tool for high throughput sequence data*. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*. 2011;**27**:1691–1692. https://doi.org/10.1093/bioinformatics/btr174

Bedre R, Mangu VR, Srivastava S, Sanchez LE, Baisakh N. Transcriptome analysis of smooth cordgrass (*Spartina alterniflora* Loisel), a monocot halophyte, reveals candidate genes involved in its adaptation to salinity. *BMC Genomics*. 2016;**17**:657. https://doi.org/10.1186/s12864-016-3017-3

Bertness MD, Ewanchuk PJ, Silliman BR. Anthropogenic modification of New England salt marsh landscapes. *Proc Natl Acad Sci*. 2002;**99**:1395–1398. https://doi.org/10.1073/pnas.022447299

Bruna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom Bioinform*. 2020;**2**:lqaa026. https://doi.org/10.1093/nargab/lqaa026

Bruna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform*. 2021;**3**:lqaa108. https://doi.org/10.1093/nargab/lqaa108

Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;**12**:59–60. https://doi.org/10.1038/nmeth.3176

Bushnell B. *BBMap: A Fast, Accurate, Splice-Aware Aligner*. Lawrence Berkeley National Laboratory. LBNL Report #: LBNL-7065E. 2014. Retrieved from https://escholarship.org/uc/item/1h3515gn

Cahoon DR, McKee KL, Morris JT. How plants influence resilience of salt marsh and mangrove wetlands to sea-level rise. *Estuar Coasts*. 2021;**44**:883–898. https://doi.org/10.1007/s12237-020-00834-w

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;**10**. https://doi.org/10.1186/1471-2105-10-421

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit—interactive quality assessment of genome assemblies. *G3 Genes|Genomes|Genetics*. 2020;**10**:1361–1374. https://doi.org/10.1534/g3.119.400908

Chapman VJ. Succession on the New England salt marshes. *Ecology*. 1940;**21**:279–282. https://doi.org/10.2307/1930502

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021;**18**:170–175. https://doi.org/10.1038/s41592-020-01056-5

Cheng H, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmell NJ, Li H. Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol*. 2022;**40**:1332–1335. https://doi.org/10.1038/s41587-022-01261-x

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;**10**. https://doi.org/10.1093/gigascience/giab008

De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics*. 2023;**39**:btad311. https://doi.org/10.1093/bioinformatics/btad311

Drake K, Halifax H, Adamowicz SC, Craft C. Carbon sequestration in tidal salt marshes of the Northeast United States. *Environ Manag*. 2015;**56**:998–1008. https://doi.org/10.1007/s00267-015-0568-z

Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostofa R, Pham M, Glenn St Hilaire B, Yao W, Stamenova E, et al. The Juicebox Assembly Tools module facilitatesde novoassembly of mammalian genomes with chromosome-length scaffolds for under $1000. 2018. https://doi.org/10.1101/254797

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst*. 2016a;**3**:99–101. https://doi.org/10.1016/j.cels.2015.07.012

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. 2016b;**3**:95–98. https://doi.org/10.1016/j.cels.2016.07.002

Emery NC, Ewanchuk PJ, Bertness MD. Competition and salt-marsh plant zonation: stress tolerators may be dominant competitors. *Ecology*. 2001;**82**:2471–2485. https://doi.org/10.1890/0012-9658(2001)082[2471:CASMPZ]2.0.CO;2

Farzi A, Borghei SM, Vossoughi M. The use of halophytic plants for salt phytoremediation in constructed wetlands. *Int J Phytoremediation*. 2017;**19**:643–650. https://doi.org/10.1080/15226514.2016.1278423

Gabriel L, Hoff KJ, Bruna T, Borodovsky M, Stanke M. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics*. 2021;**22**:566. https://doi.org/10.1186/s12859-021-04482-0

Gotoh O. A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. *Nucleic Acids Res*. 2008;**36**:2630–2638. https://doi.org/10.1093/nar/gkn105

Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. *Bioinformatics*. 2014;**30**:2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*. 2016;**32**:767–769. https://doi.org/10.1093/bioinformatics/btv661

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. In: Kollmar M, editor. *Gene prediction*. Methods in Molecular Biology, vol 1962. New York, NY: Humana, 2019, 65–95. https://doi.org/10.1007/978-1-4939-9173-0_5

Iwata H, Gotoh O. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Res*. 2012;**40**:e161–e161. https://doi.org/10.1093/nar/gks708

Jia K-H, Wang Z-X, Wang L, Li G-Y, Zhang W, Wang X-L, Xu F-J, Jiao S-Q, Zhou S-S, Liu H, et al. SubPhaser: a robust allopolyploid subgenome phasing method based on subgenome-specific k-mers. *New Phytol*. 2022;**235**:801–809. https://doi.org/10.1111/nph.18173

Kadereit G, Ball P, Beer S, Mucina L, Sokoloff D, Teege P, Yaprak AE, Freitag H. A taxonomic nightmare comes true: phylogeny and biogeography of glassworts (*Salicornia* L., Chenopodiaceae). *Taxon*. 2007;**56**:1143–1170. https://doi.org/10.2307/25065909

Kokot M, Długosz M, Deorowicz S. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics*. 2017;**33**:2759–2761. https://doi.org/10.1093/bioinformatics/btx304

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019;**37**:540–546. https://doi.org/10.1038/s41587-019-0072-8

Krause JR, Watson EB, Wigand C, Maher N. Are tidal salt marshes exposed to nutrient pollution more vulnerable to sea level rise? *Wetlands*. 2020;**40**:1539–1548. https://doi.org/10.1007/s13157-019-01254-8

Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics*. 2021;**37**:4572–4574. https://doi.org/10.1093/bioinformatics/btab705

Li FW, Harkess A. A guide to sequence your favorite plant genomes. *Appl Plant Sci*. 2018;**6**:e1030. https://doi.org/10.1002/aps3.1030

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;**25**:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res*. 2005;**33**:6494–6506. https://doi.org/10.1093/nar/gki937

Lomsadze A, Burns PD, Borodovsky M. Integration of mapped RNA-seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res*. 2014;**42**:e119–e119. https://doi.org/10.1093/nar/gku557

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet*. 2011;**17**:10–12. https://doi.org/10.14806/ej.17.1.200

Mian S, Christenhusz MJM, Leitch IJ, Leitch AR, Royal Botanic Gardens Kew Genome Acquisition Lab, Plant Genome Sizing collective, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team, Wellcome Sanger Institute Scientific Operations: Sequencing Operations, Wellcome Sanger Institute Tree of Life Core Informatics team, et al. The genome sequence of purple glasswort, *Salicornia ramosissima* woods (Amaranthaceae) [version 1; peer review: 1 approved, 2 approved with reservations]. *Wellcome Open Res*. 2024;**9**:257. https://doi.org/10.12688/wellcomeopenres.21552.1

Milic D, Lukovic J, Ninkov J, Zeremski-Skoric T, Zoric L, Vasin J, Milic S. Heavy metal content in halophytic plants from inland and maritime saline areas. *Cent Eur J Biol*. 2012;**7**:307–317. https://doi.org/10.2478/s11535-012-0015-6

Mukherjee P, Pramanick P, Zaman S, Mitra A. Phytoremediation of heavy metals by the dominant mangrove associate species of INDIAN SUNDARBANS. *J Environ Eng Landsc Manag*. 2021;**29**:391–402. https://doi.org/10.3846/jeelm.2021.15773

Oxford Nanopore Technologies. 2023. modkit (0.3.1): GitHub; [Source code]. https://github.com/nanoporetech/modkit

Oxford Nanopore Technologies. 2024. Dorado (0.5.1): GitHub; [Source code]. https://github.com/nanoporetech/dorado

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;**26**:841–842. https://doi.org/10.1093/bioinformatics/btq033

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;**11**:1432. https://doi.org/10.1038/s41467-020-14998-3

Rey E, Maughan PJ, Maumus F, Lewis D, Wilson L, Fuller J, Schmöckel SM, Jellen EN, Tester M, Jarvis DE. A chromosome-scale assembly of the quinoa genome provides insights into the structure and dynamics of its subgenomes. *Commun Biol*. 2023;**6**:1263. https://doi.org/10.1038/s42003-023-05613-4

Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020;**21**:245. https://doi.org/10.1186/s13059-020-02134-9

Salazar OR, Chen K, Melino VJ, Reddy MP, Hřibová E, Čížková J, Beránková D, Arciniegas Vega JP, Cáceres Leal LM, Aranda M, et al. SOS1 tonoplast neo-localization and the RGG protein SALTY are important in the extreme salinity tolerance of *Salicornia bigelovii*. *Nat Commun*. 2024;**15**:4279. https://doi.org/10.1038/s41467-024-48595-5

Sharma A, Gontia I, Agarwal PK, Jha B. Accumulation of heavy metals and its biochemical responses in *Salicornia brachiata*, an extreme halophyte. *Mar Biol Res*. 2010;**6**:511–518 Article Pii 923874022. https://doi.org/10.1080/17451000903434064

Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res*. 2019;**47**:W65–W73. https://doi.org/10.1093/nar/gkz345

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;**31**:3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smit AFA, Hubley R, Green P. *RepeatMasker Open-4.0*. 2013-2015. http://www.repeatmasker.org

Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*. 2006;**7**:62. https://doi.org/10.1186/1471-2105-7-62

Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*. 2008;**24**:637–644. https://doi.org/10.1093/bioinformatics/btn013

Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res*. 2017;**27**:737–746. https://doi.org/10.1101/gr.214270.116

Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, Wyres KL, Holt KE. Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biology*. 2021;**22**. https://doi.org/10.1186/s13059-021-02483-z

Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol*. 2017;**13**:e1005595. https://doi.org/10.1371/journal.pcbi.1005595

Winter D. pafr: Read, Manipulate and Visualize 'Pairwise mApping Format' Data. R package version 0.0.2, commit 115d2b1be6cdab5679a81a536200a03cef1ba0e9. 2025.

Wolff SL, Jefferies RL. Morphological and isozyme variation in Salicornia europaea (s.l.) (Chenopodiaceae) in northeastern North America. *Canadian Journal of Botany*. 1987;**65**:1410–1419. https://doi.org/10.1139/b87-195

Zeng X, Yi Z, Zhang X, Du Y, Li Y, Zhou Z, Chen S, Zhao H, Yang S, Wang Y, Chen G. *Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes*. 2023. *bioRxiv*, 2023.2011.2018.567668. https://doi.org/10.1101/2023.11.18.567668

Zhang FG, Xiao X, Xu K, Cheng X, Xie T, Hu JH, Wu XM. Genome-wide association study (GWAS) reveals genetic loci of lead (Pb) tolerance during seedling establishment in rapeseed (*Brassica napus* L.). *BMC Genomics*. 2020;**21**. https://doi.org/10.1186/s12864-020-6558-4

Journal of Heredity, 2026, 117, 357–365
https://doi.org/10.1093/jhered/esaf062
Genome Resources