GBE

Genome Assembly of *Spinifex sericeus* Provides Insights into Repetitive Content and Haplotype Variation in a Dioecious Coastal Grass

Stephanie H. Chen [1], Jason G. Bragg [2], Ashley Jones [3,*]

[1]Centre for Australian National Biodiversity Research (a Joint Venture Between Parks Australia and CSIRO), Canberra, ACT 2601, Australia

[2]Research Centre for Ecosystem Resilience, Botanic Gardens of Sydney, Sydney, NSW 2000, Australia

[3]Research School of Biology, Australian National University, Canberra, ACT 2601, Australia

*Corresponding author: E-mail: ashley.jones@anu.edu.au.

Accepted: February 18, 2026

Abstract

Spinifex sericeus (beach spinifex) is a perennial, dioecious grass native to the coastal ecosystem of Oceania that has an important role in dune stability. We present a near telomere-to-telomere, haplotype-resolved nuclear genome and complete organelle genomes of *S. sericeus*. The nuclear genome is highly repeat-rich, dominated by long terminal repeat (LTR) retrotransposons. Comparative analysis of the two haplotypes reveals extensive structural variation and high levels of duplication that are suggestive of active and expanding LTRs that may drive genome evolution and expansion in this species. The chloroplast genome exhibits heteroplasmy, having two distinct circular configurations spanning 138 kbp each. The complex mitochondrial genome consists of a large linear component spanning 1.85 Mbp and a small circular 134 kbp component. These genomic resources provide a foundation for advancing research on sex determination and stress adaptation in grasses, as well as practical applications in crop improvement and genetically informed coastal dune restoration.

Key words: genome assembly, Poaceae, spinifex, telomere-to-telomere, structural variants.

Significance

The reference genome of *Spinifex sericeus* provides a foundational resource for understanding adaptations to the harsh coastal dune environment, including saline conditions, extreme temperatures, and drought. The genome enables comparative studies within Poaceae, where current reference genomes are dominated by crop species, and supports conservation and restoration efforts for coastal ecosystem management.

Introduction

Grasses (Poaceae) represent one of the largest plant families, comprising c. 11,800 species across 791 genera (Soreng et al. 2022). Poaceae dominate terrestrial ecosystems and include the world's most important cereal crops, positioning them as both an ecologically and economically critical lineage for genomic research. Despite the challenges posed by their characteristically large and repetitive genomes, Poaceae has emerged as a model clade in plant genomics (Buell 2009). Advances in long-read sequencing technology, coupled with decreasing costs, have significantly accelerated

research in Poaceae genomics, making high-quality telomere-to-telomere (T2T) genomes accessible (Xie et al. 2024). However, sequencing efforts have primarily focused on agriculturally important species such as rice, maize, wheat, and sorghum. For example, rice (*Oryza sativa* L.) was the first crop sequenced (International Rice Genome Sequencing Project 2005), leading to improvements in crop breeding and production. Wild grass relatives remain a largely unexplored genomic resource, having evolved adaptive mechanisms to abiotic stresses that are increasingly valuable for crop improvement under climate change (Hultgren et al. 2025).

Spinifex sericeus is a dioecious perennial grass occurring on coastal foredunes in its native range of eastern Australia, New Zealand, New Caledonia, and Tonga. As a pioneer dune species, this grass survives in hostile environments (Maze and Whalley 1992) through adaptations to high salinity and heat but faces potential threats from human disturbance, grazing (Ramsey and Engeman 1994), and diseases such as smut (Kirby 1988). *Spinifex sericeus* stabilizes the frontal dune area and is valuable for dune restoration, improving both the landscape and ecological integrity in coastal regions (Jenks and Brake 2001).

Sequencing the genome of *S. sericeus* will enhance our understanding of *Spinifex* biology and expand the genomic resources available for Poaceae beyond crop species. In this study, we present the first assembly and annotation of a near T2T complete haplotype-resolved nuclear genome for *S. sericeus* and the complete chloroplast and mitochondrial genomes, using PacBio HiFi long-read sequencing data, Oxford Nanopore Technologies (ONT) ultra-long reads, and Hi-C data.

Results and Discussion

We sampled a *S. sericeus* plant in its native range in Australia (Fig. 1a, b). Sequencing yielded 91.7 Gbp PacBio HiFi reads, 192.1 Gbp ONT reads, and 76.2 Gbp Hi-C reads (Table S1). All HiFi reads were used as input into assembly (46× coverage), whereas ONT reads were filtered for ≥20 kbp and ≥ Q10, resulting in 151.9 Gbp of input reads (76× coverage). After Hi-C read trimming and quality filtering (≥20 bp and ≥ Q20), there were 74.5 Gbp of data (37× coverage). Using hifiasm (UL), the initial de novo assemblies had 490 contigs (haplotype 1) and 122 contigs (haplotype 2) with N50 of 195 and 186 Mbp, respectively (Table S2 and Fig. S1). Hi-C scaffolding with YaHS made 13 joins, reducing the number of contigs to 485 for haplotype 1 and 111 for haplotype 2; N50 increased to 222 and 233 Mbp, respectively (Table S3). The scaffolding results also validated a hifiasm (UL) test run which incorporated Hi-C data via the --dual-scaf option.

We generated a near T2T complete haplotype-resolved nuclear genome for *S. sericeus*, with each haplotype comprising nine chromosome-length scaffolds (Table 1 and Fig. 1c). The species has a 2C value of 5.41 pg with $2x = 18$ (Murray et al. 2005), which approximates the haploid genome size at 2.6 Gbp. Our assembly was smaller at 1.97 Gbp for haplotype 1 and 2.04 Gbp for haplotype 2, and we confirmed diploidy via Smudgeplot (Fig. S2).

Both haplotypes had very high BUSCO genome completeness scores of 99.7%. For haplotype 1, 97.44% was single-copy, with 2.26% duplicated. Similarly, for haplotype 2, 97.48% was single-copy, and 2.20% duplicated. Both haplotypes were independently annotated and contained approximately 47,000 protein-coding genes. Overall proteome completeness exceeded 99% for both haplotypes; however we observed higher rates of duplication. For haplotype 1, 84.08% was single-copy and 15.07% duplicated, whereas for haplotype 2, 83.95% and 15.14% were single-copy and duplicated, respectively. Our proteome results were comparable to the best RefSeq Poaceae genomes available, where single-copy scores ranged from 37.36% to 86.18%, reflecting the relatively high proportion of duplicated genes that appears characteristic of Poaceae (Table S4). Haplotype 1 contained a notably higher number of rRNAs compared to haplotype 2 (2,409 vs 86), while tRNA numbers were identical at 485 (Table 1). The smaller number of rRNAs annotated on haplotype 2 may reflect intraspecific variation, but it is also possible that a cluster of rRNAs on haplotype 2 were not successfully assembled. We note that haplotype 1 contains rRNAs on chromosome 8 (96 to 98 Mbp) near an estimated 2.65 Mbp assembly gap on haplotype 2 (102 to 105 Mbp; Fig. 1d).

The *S. sericeus* genome contained a high proportion of repeats at over 82%, dominated by long terminal repeat (LTR) retrotransposons, notably Gypsy and Copia elements (Table S5); this number was higher than the estimate given by GenomeScope using both HiFi (57%) and ONT data (57%) (Fig. S3). Using both tidk (Fig. S4) and seqtk (Table S6), 31 out of 36 telomeres (motif AAACCCT) were identified (summarized in Table S7). Haplotype 1 consisted of 5/9 T2T scaffolds, and haplotype 2 consisted of 8/9 T2T scaffolds; there are only 13 gaps in total (Table S8). The *S. sericeus* genome had a high assembly quality value (QV) of 62 for each haplotype, indicating a base accuracy of >99.9999%, which exceeds human T2T genome projects (Cheng et al. 2026). This also highlights how long-read sequencing accuracy and de novo assembly for nonmodel species have improved substantially (Ferguson et al. 2022).

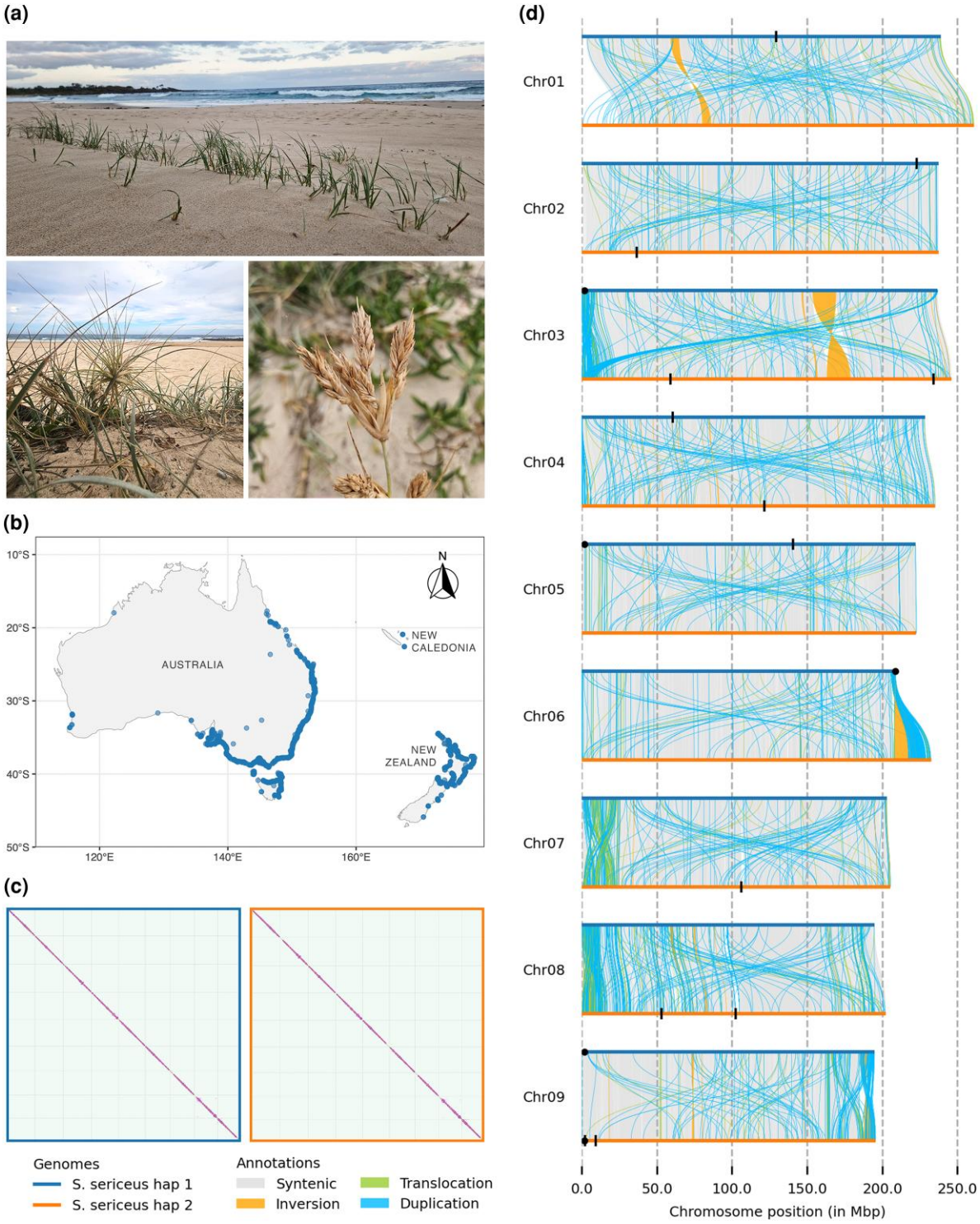

Fig. 1. a) *Spinifex sericeus* growing at Kioloa, New South Wales, Australia (top; photograph by A.J.) and close up of flowers at Maroubra, New South Wales (female on bottom left and male on bottom right; photographs by J.G.B.). b) Map of global occurrence records from the Biodiversity Information Facility (GBIF.org 2025). c) Hi-C contact maps of haplotype 1 (left) and haplotype 2 (right) visualized in PretextView with gray lines separating scaffolds. d) Synteny and structural variation between the two haplotypes of the *S. sericeus* genome. Haplotypes 1 and 2 were aligned, and genomic regions were classified as syntenic, inverted, translocated, duplicated, or haplotype specific (remaining white). Only structural variants ≥10 kbp are visualized. Vertical black lines on the genome represent gaps, where scaffolding occurred (13 gaps total). Circular black markers represent missing telomeres (31/36 present).

Table 1 Summary of the *S. sericeus* genome assembly, annotation, and quality assessment. Karyotype is $2n = 18$

Statistic	Haplotype 1	Haplotype 2
Total length (bp)	1,966,204,384	2,039,249,938
Scaffolds	9	9
N50 (bp)	222,255,665	232,667,679
Telomeres	14/18	17/18
T2T contigs	2/9	3/9
T2T scaffolds	5/9	8/9
Gaps	4	9
GC (%)	46.35	46.43
BUSCO genome completeness (poales_odb12, $n = 6,282$)	99.7%	99.7%
Single-copy	97.44%	97.48%
Duplicated	2.26%	2.20%
Fragmented	0.03%	0.03%
Missing	0.27%	0.29%
Protein-coding genes (GeMoMa)	47,411	47,342
rRNAs	2,409	86
tRNAs	485	485
BUSCO proteome completeness (poales_odb12, $n = 6,282$)	99.2%	99.1%
Single-copy	84.08%	83.95%
Duplicated	15.07%	15.14%
Fragmented	0.22%	0.27%
Missing	0.62%	0.64%
Repeats (%)	82.06	82.30
QV	61.79	62.33

To assess the conservation of genome structure and identify structural variations, the two haplotypes were aligned and compared. Both haplotypes were highly syntenic, with 92% of haplotype 1 (1,846 Mbp) and 89% of haplotype 2 (1,847 Mbp) showing conserved sequence arrangement (Fig. 1d and Table S9). The remaining sequences contained substantial structural variations distributed across the genome. Duplications were the most prominent structural variant, comprising 8,517 events in haplotype 1 totaling 58 Mbp (2.89%) and 9,318 events in haplotype 2 totaling 75 Mbp (3.60%). We observed duplications genome-wide, including inter-haplotype events spanning entire chromosome arms, potentially reflecting underlying genome organization. The majority of duplications ranged between 5 and 20 kbp, consistent with the sizes of annotated LTR retrotransposons including Gypsy and Copia elements, suggesting that transposable element activity is a key driver of genomic structural variation. This prevalence of duplications aligns with emerging evidence that they serve as a critical evolutionary mechanism in Poaceae, enabling rapid adaptation to diverse environmental landscapes (Ahrens et al. 2020; Bellec et al. 2023).

Haplotype-specific sequences represented 35 Mbp (1.72%) in haplotype 1 and 71 Mbp (3.40%) in haplotype 2, comprising novel insertions, deletions, and highly diverged regions. These diverged regions may also include duplications that have accumulated mutations beyond recognition (Ferguson et al. 2024). Translocations were moderately abundant, with 1,412 events spanning approximately 42 Mbp (2.08%) in haplotype 1 and 41 Mbp (1.97%) in haplotype 2. Inversions were rare but large, with only 95 events detected that cumulatively affected 35 Mbp (1.75%) of haplotype 1 and 46 Mbp (2.23%) of haplotype 2. Read mapping to both haplotypes and subsequent coverage analysis verified the structural variants and haplotype-specific regions and confirmed high confidence in the genome assembly quality (Fig. S5 and Table S10). For the PacBio HiFi reads, 98.62% mapped to the nuclear genome, 1.06% were organelle reads, and only 0.32% remained unmapped. Similarly, for ONT ultra-long reads, 99.45% mapped to the nuclear genome, 0.33% were organelle reads, and 0.22% remained unmapped. The extensive structural variation between haplotypes underscores the remarkable genome evolution within this species and suggests that structural variants contribute to genomic diversity and adaptive potential.

We also successfully assembled complete chloroplast and mitochondrial genomes, revealing complex structural arrangements in both organelles. The *S. sericeus* chloroplast exhibited heteroplasmy, having two distinct circular configurations of 137,659 bp each, which is consistent with other Poales grasses (Zhou et al. 2025). The mitochondrial genome consisted of two components: a considerably large linear sequence of 1,851,594 bp and a circular sequence of 134,119 bp (Fig. S6). It is possible the linear sequence may be incomplete; however, mitochondrial genomes consisting of a linear and circular component have been reported, for example, *Alnus glutinosa* (Zhou et al. 2025). Similarly, Poales mitochondria genomes have been observed to be larger on average, for example, the 2.73 Mbp mitochondria of *Carex laevigata* (Zhou et al. 2025).

This reference genome of *S. sericeus* represents a foundational resource for understanding grass genome evolution and adaptive capacity under abiotic stress. Thriving under the high salinity, temperature extremes, and intense light of coastal dunes, *S. sericeus* offers opportunities to identify genetic mechanisms underlying stress tolerance, which are critical for developing climate-resilient crops (Melino and Tester 2023). As a dioecious species, this genome also enables investigation of sex-biased gene expression across reproductive and vegetative tissues, such as inflorescences, leaves, and roots (Sanderson et al. 2019). Field studies have observed a male bias in the observed ramet sex ratio (Maze and Whalley 1990), raising questions about the underlying mechanisms driving this imbalance. Genes

involved in sex determination and ecological adaptations could be investigated and contribute to genetically informed restoration strategies to enhance coastal dune resilience and conservation.

Materials and Methods

Sampling, DNA Extraction, and Sequencing

Cuttings were sampled from a single plant in Yarra Bay Beach, New South Wales, Australia (33.9774° S, 151.2280° E; BioSample SAMN41648390) and propagated in a growth chamber in the original sand substrate supplemented with a seaweed solution (Seasol) fortnightly. The sex of the reference genome is currently unknown as the plant was not flowering at the time of collection, nor has it flowered in the 1.5 years that it has been in cultivation. Young leaf blades were harvested and cryogenically stored at −80 °C for DNA extractions and Hi-C.

High-molecular-weight DNA was extracted using a magnetic bead-based protocol as previously described (Jones et al. 2021). The DNA quality was assessed on a Femto Pulse system (Fig. S7) using a genomic DNA 165 kbp kit (Agilent Technologies) and subsequently size selected for fragments ≥20 kbp using a BluePippin (Sage Science). A Pacific Biosciences (PacBio) DNA library was prepared using the SMRTbell prep kit 3.0 and sequenced on the Revio platform using a 25 M SMRT cell, to generate highly accurate HiFi reads (≥Q20). Two Oxford Nanopore Technologies (ONT) native DNA libraries were prepared using the Ligation Sequencing Kit V14 (SQK-LSK114) and sequenced on the PromethION P24 platform using two FLO-PRO114 M R10.4.1 flow cells, to generate ultra-long reads. The ONT flow cells were washed, re-primed, and reloaded twice when sequencing declined (ONT Flow Cell Wash Kit EXP-WSH004). ONT sequencing data (pod5) was basecalled to fastq with ONT Dorado v7.4.12 using the super-accurate DNA model (sup v4.3.0, e8.2, 400 bps). The ONT fastq reads were filtered for ≥20 kbp and ≥ Q10 using chopper v0.8.0 (De Coster et al. 2018).

A chromosome conformation Hi-C library was prepared using a Phase Genomics Proximo Hi-C Plant Kit v4 (KT3040B) and sequenced on an Illumina NovaSeq 6000 using an S4 flow cell with a 300-cycle kit (150 bp paired-end sequencing). Hi-C reads were trimmed for Illumina adapters (–nextera) and filtered for ≥20 bp and ≥ Q20. Read pairs were validated with Trim Galore v0.6.2 (Krueger et al. 2023).

Genome Assembly

Independently using PacBio HiFi reads and filtered ONT ultra-long reads, the nuclear genome was profiled by k-mer frequencies ($k = 31$) to assess genome size, heterozygosity, and ploidy, with FastK v1.1, GenomeScope v2.0.1, and Smudgeplot v0.4.0 (Ranallo-Benavidez et al. 2020). A de novo haplotype-resolved genome assembly was performed with hifiasm (UL) v0.24.0-r703 (Cheng et al. 2024), using PacBio HiFi reads, filtered ONT ultra-long reads (–ul), filtered Illumina Hi-C reads (–h1 --h2), and specifying the representative Poaceae telomere motif (–telo-m AAACCCT). To scaffold the genome, Hi-C reads were mapped with HiC-Pro v3.1.0 (Servant et al. 2015) and scaffolding was performed with YaHS v1.2.2 (Zhou et al. 2023). Hi-C contact maps were generated using PretextMap v0.1.9 (Harry 2022) and visualized with PretextView v1.0.5 (Harry and Guan 2025). Gaps of unknown length were standardized to 100 bp. Scaffolds less than 1 Mbp were filtered out so that the remaining sequences were the nine nuclear chromosomes. The nine chromosomes were ordered by length, according to haplotype 1.

To assemble organelle genomes, the PacBio HiFi reads were first randomly subsampled to 2 M reads using seqtk sample v1.5 (r133) (Li 2025), capturing hundreds to thousands of times coverage of each organelle genome. The chloroplast genome was assembled with TIPPo v1.3.0 (Xian et al. 2025), and the mitochondria genome was assembled with OatK v1.0 (Zhou et al. 2025).

Assembly Quality Assessment

All genome assembly graphs were inspected with Bandage v0.8.1 (Wick et al. 2015). Nuclear genome completeness was assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO) with the Poales OrthoDB v12 database ($n = 6{,}282$), using compleasm v0.2.7 (Huang and Li 2023). To assess base accuracy of the genome, yak v0.1 (r56) was used to compare k-mer spectra ($k = 31$) between the assembly and PacBio HiFi reads to calculate a coverage-adjusted assembly quality value (QV) (Cheng et al. 2021). To assess genome coverage and mapping statistics, PacBio HiFi reads and filtered ONT ultra-long reads were independently aligned to the genome using Minimap2 v2.29 (Li 2018). Coverage was plotted using the JVarkit v2024.08.25 tool WGSCoveragePlotter (Lindenbaum 2024).

Genome Annotation

Genes were annotated with GeMoMa v1.9 (Keilwagen et al. 2019) with five RefSeq genomes: rice (*Oryza sativa*; GCF_034140825.1) (Shang et al. 2023), corn (*Zea mays*; GCF_902167145.1) (Hufford et al. 2021), common reed (*Phragmites australis*; GCF_958298935.1) (Christenhusz

et al. 2024), green foxtail (*Setaria viridis*; GCF_005286985.2) (Thielen et al. 2020), and *Arabidopsis thaliana* (TAIR 10.1; GCF_000001735.4) (Lamesch et al. 2012). The longest isoform per locus was extracted using AGAT v1.0.0 (Dainat 2022). Completeness was assessed with compleasm v0.2.7 (Huang and Li 2023) in protein mode with the Poales OrthoDB v12 database. For comparison, the four RefSeq Poaceae genomes used for annotation were also assessed with compleasm after the same isoform filtering.

The rRNA genes were predicted with Barrnap v0.9 (Seemann 2018), and tRNAs were predicted with tRNAscan-SE v2.0.12 (Lowe and Chan 2016), applying the high confidence filter. Repeats were annotated with a custom library using RepeatModeler v2.0.2a with -engine ncbi (Flynn et al. 2020) and RepeatMasker v4.1.2 (Tarailo-Graovac and Chen 2009). Telomeric repeats were identified with tidk v0.2.65 (Brown et al. 2025) and seqtk v1.5 (r133) (Li 2025).

Synteny and Structural Variation Between Haplotypes

To assess nuclear genomic differences, the two haplotypes were aligned to each other with the MUMmer4 v4.0.1 tool nucmer (–maxmatch -l 50 -b 500 -c 200) (Marçais et al. 2018). All shared 50-mers between haplotypes were identified, and adjacent 50-mers were merged into a single alignment. Alignments were filtered for ≥200 bp and ≥95% sequence identity with the MUMmer4 tool delta-filter. Synteny and genomic rearrangements were identified with SyRI v1.6.3 (Goel et al. 2019), and results ≥10 kbp were visualized with plotsr v1.1.1 (Goel and Schneeberger 2022).

Supplementary Material

Supplementary material is available at Genome Biology and Evolution online.

Acknowledgments

We acknowledge Australia's First Nations Peoples as the Traditional Custodians of the land and waters on which this work was conducted and pay our respects to all Elders past and present. We would like to thank the Biomolecular Resource Facility at the John Curtin School of Medical Research, ANU in Canberra, Australia, where sequencing was conducted. This research acknowledges the support provided by NCRIS-enabled Bioplatforms Australia infrastructure. This research was undertaken with the assistance of resources from the National Computational Infrastructure (NCI Australia), an NCRIS-enabled capability supported by the Australian Government.

Author Contributions

Stephanie H. Chen (Formal Analysis, Investigation, Methodology, Software, Visualization, Writing—Original draft, Writing—review & editing), Jason G. Bragg (Investigation, Writing—review & editing), Ashley Jones (Conceptualization, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing—original draft, Writing—review & editing)

Funding

This research was funded by a Centre for Biodiversity Analysis (CBA) Ignition Grant awarded to A.J. and S.H.C. A.J. was supported by the Australian Government through an Australian Research Council (ARC) Discovery Early Career Researcher Award (DECRA) DE260100171.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

The *S. sericeus* reference genome and raw sequencing data (PacBio HiFi, ONT, and Hi-C data) were deposited to NCBI under BioProject PRJNA1119262 and PRJNA1346877. Additional data supporting this work, including genome annotation files, are available on the CSIRO Data Access Portal (https://doi.org/10.25919/eh42-6453). All raw sequencing data and assembled genomes have been shared with the broader public via appropriate biological databases.

Literature Cited

Ahrens CW, et al. Spatial, climate and ploidy factors drive genomic diversity and resilience in the widespread grass *Themeda triandra*. Mol Ecol. 2020:29:3872–3888. https://doi.org/10.1111/mec.15614.

Bellec A, et al. Tracing 100 million years of grass genome evolutionary plasticity. Plant J. 2023:114:1243–1266. https://doi.org/10.1111/tpj.16185.

Brown MR, Manuel Gonzalez de La Rosa P, Blaxter M. tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. Bioinformatics. 2025:41:btaf049. https://doi.org/10.1093/bioinformatics/btaf049.

Buell CR. Poaceae genomes: going from unattainable to becoming a model clade for comparative plant genomics. Plant Physiol. 2009:149:111–116. https://doi.org/10.1104/pp.108.128926.

Cheng H, et al. 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, et al. Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. Nat Methods. 2024:21:967–970. https://doi.org/10.1038/s41592-024-02269-8.
Cheng H, et al. Efficient near-telomere-to-telomere assembly of nanopore simplex reads. Nature. 2026. https://doi.org/10.1038/s41586-026-10105-6.
Christenhusz MJM, et al. The genome sequence of common reed, *Phragmites australis* (Cav.) Steud. (Poaceae). Wellcome Open Res. 2024:9:577. https://doi.org/10.12688/wellcomeopenres.23143.1.
Dainat J. 2022. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format. https://github.com/NBISweden/AGAT
De Coster W, et al. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018:34:2666–2669. https://doi.org/10.1093/bioinformatics/bty149.
Ferguson S, et al. Species-specific basecallers improve actual accuracy of nanopore sequencing in plants. Plant Methods. 2022:18:137. https://doi.org/10.1186/s13007-022-00971-2.
Ferguson S, et al. Plant genome evolution in the genus *Eucalyptus* is driven by structural rearrangements that promote sequence divergence. Genome Res. 2024:34:606–619. https://doi.org/10.1101/gr.277999.123.
Flynn JM, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci. 2020:117:9451–9457. https://doi.org/10.1073/pnas.1921046117.
GBIF.org. 2025. Occurrence download. https://www.gbif.org/occurrence/download/0007018-251120083545085. https://doi.org/10.15468/DL.JVVX77.
Goel M, Schneeberger K. plotsr: visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics. 2022:38:2922–2926. https://doi.org/10.1093/bioinformatics/btac196.
Goel M, Sun H, Jiao W-B, Schneeberger K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 2019:20:277. https://doi.org/10.1186/s13059-019-1911-0.
Harry E. 2022. PretextMap. https://github.com/sanger-tol/PretextMap
Harry E, Guan S. 2025. PretextView. https://github.com/sanger-tol/PretextView
Huang N, Li H. compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics. 2023:39:btad595. https://doi.org/10.1093/bioinformatics/btad595.
Hufford MB, et al. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. Science. 2021:373:655–662. https://doi.org/10.1126/science.abg5289.
Hultgren A, et al. Impacts of climate change on global agriculture accounting for adaptation. Nature. 2025:642:644–652. https://doi.org/10.1038/s41586-025-09085-w.
International Rice Genome Sequencing Project. The map-based sequence of the rice genome. Nature. 2005:436:793–800. https://doi.org/10.1038/nature03895.
Jenks GK, Brake LA. Enhancing dune function and landscape integrity using active and passive bio-engineering, Bay of Plenty coast, New Zealand. J Coast Res. 2001:528–534. https://www.jstor.org/stable/25736318.
Jones A, et al. High-molecular weight DNA extraction, clean-up and size selection for long-read sequencing. PLoS One. 2021:16:e0253830. https://doi.org/10.1371/journal.pone.0253830.
Keilwagen J, Hartung F, Grau J. GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. Methods Mol Biol Clifton NJ. 2019:1962:161–177. https://doi.org/10.1007/978-1-4939-9173-0_9.
Kirby GC. The population biology of a smut fungus, *Ustilago spinificis* Ludw. I. Geographic distribution and abundance. Aust J Bot. 1988:36:339–346. https://doi.org/10.1071/bt9880339.
Krueger F, et al. FelixKrueger/TrimGalore. 2023. [accessed 2025 Dec 2]. https://zenodo.org/records/7598955. https://doi.org/10.5281/zenodo.7598955.
Lamesch P, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 2012:40:D1202–D1210. https://doi.org/10.1093/nar/gkr1090.
Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018:34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
Li H. 2025. seqtk. https://github.com/lh3/seqtk
Lindenbaum P. 2024. JVarkit. https://github.com/lindenb/jvarkit
Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 2016:44:W54–W57. https://doi.org/10.1093/nar/gkw413.
Marçais G, et al. MUMmer4: a fast and versatile genome alignment system. PLOS Comput Biol. 2018:14:e1005944. https://doi.org/10.1371/journal.pcbi.1005944.
Maze KM, Whalley RDB. Sex-ratios and related characteristics in *Spinifex sericeus* (Poaceae). Aust J Bot. 1990:38:153–160. https://doi.org/10.1071/bt9900153.
Maze KM, Whalley RDB. Germination, seedling occurrence and seedling survival of *Spinifex sericeus* R. Br. (Poaceae). Aust J Ecol. 1992:17:189–194. https://doi.org/10.1111/j.1442-9993.1992.tb00797.x.
Melino V, Tester M. Salt-tolerant crops: time to deliver. Annu Rev Plant Biol. 2023:74:671–696. https://doi.org/10.1146/annurev-arplant-061422-104322.
Murray BG, De Lange PJ, Ferguson AR. Nuclear DNA variation, chromosome numbers and polyploidy in the endemic and indigenous grass flora of New Zealand. Ann Bot. 2005:96:1293–1305. https://doi.org/10.1093/aob/mci281.
Ramsey DSL, Engeman RM. Patterns of grazing on coastal dune systems by insular populations of two species of macropod. Wildl Res. 1994:21:107–113. https://doi.org/10.1071/wr9940107.
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.
Sanderson BJ, et al. Sex-biased gene expression in flowers, but not leaves, reveals secondary sexual dimorphism in *Populus balsamifera*. New Phytol. 2019:221:527–539. https://doi.org/10.1111/nph.15421.
Seemann T. 2018. barrnap. https://github.com/tseemann/barrnap
Servant N, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 2015:16:259. https://doi.org/10.1186/s13059-015-0831-x.
Shang L, et al. A complete assembly of the rice Nipponbare reference genome. Mol Plant. 2023:16:1232–1236. https://doi.org/10.1016/j.molp.2023.08.003.
Soreng RJ, et al. A worldwide phylogenetic classification of the Poaceae (Gramineae) III: an update. J Syst Evol. 2022:60:476–521. https://doi.org/10.1111/jse.12847.
Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinforma. 2009:25:4.10.1–4.10.14. https://doi.org/10.1002/0471250953.bi0410s25.
Thielen PM, et al. Reference genome for the highly transformable *Setaria viridis* ME034V. G3 Bethesda Md. 2020:10:3467–3478. https://doi.org/10.1534/g3.120.401345.
Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics.

2015:31:3350–3352. https://doi.org/10.1093/bioinformatics/btv383.
Xian W, et al. TIPPo: a user-friendly tool for de novo assembly of organellar genomes with high-fidelity data. Mol Biol Evol. 2025:42:msae247. https://doi.org/10.1093/molbev/msae247.
Xie L, et al. Technology-enabled great leap in deciphering plant genomes. Nat Plants. 2024:10:551–566. https://doi.org/10.1038/s41477-024-01655-6.
Zhou C, et al. Oatk: a de novo assembly tool for complex plant organelle genomes. Genome Biol. 2025:26:235. https://doi.org/10.1186/s13059-025-03676-6.
Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. Bioinformatics. 2023:39:btac808. https://doi.org/10.1093/bioinformatics/btac808.

Associate editor: Vincent Castric