Scientific Data

https://doi.org/10.1038/s41597-026-06911-2

Article in Press

The chromosomal-level genome assembly and annotation of *Phyllospadix iwatensis* (Surfgrass)

Received: 28 July 2025

Accepted: 17 February 2026

Cite this article as: Wang, J., Wang, D., Zhao, K. *et al.* The chromosomal-level genome assembly and annotation of *Phyllospadix iwatensis* (Surfgrass). *Sci Data* (2026). https://doi.org/10.1038/s41597-026-06911-2

Junyi Wang, Dawei Wang, Ke Zhao, Zhining Liu & Quansheng Zhang

Title

The chromosomal-level genome assembly and annotation of *Phyllospadix iwatensis*（Surfgrass）

Junyi Wang[1], Dawei Wang[2], Ke Zhao[2], Zhining Liu[2,3]✉ & Quansheng Zhang[1,3]✉

[1]Ocean School, Yantai University, Yantai, 264005, China. [2]No. 6 Geological Team, Shandong Provincial Bureau of Geology and Mineral Resources, Weihai, 264209, China. [3]These authors jointly supervised this work and both serve as corresponding authors: Zhining Liu, Quansheng Zhang. ✉e-mail: 93444623@qq.com; ytuqsz@hotmail.com.

Abstract

Phyllospadix iwatensis is a unique seagrass species adapted to rocky substrate anchorage and dioecy and belongs to marine submerged flowering plants with a distinctive evolutionary history. The chromosomal-scale genome was constructed by integrating Illumina, PacBio HiFi, and high-throughput chromosome conformation capture (Hi-C) sequencing techniques. A total of 340.56 Mb of sequences were anchored to 10 chromosomes with an anchoring rate of 96.44%. The contig and scaffold N50 values reached 30.64 Mb and 33.59 Mb, respectively. Precisely 94.64% of the 23,198 predicted protein-coding genes received functional annotation. In the meantime, 180.19 Mb of repetitive sequences were found, representing 52.91% of the assembled genome. The chromosomal-level genome data of *P. iwatensis* will reveal its special process of differentiation and enrich the understanding of the multiple adaptations of seagrass populations to marine habitats.

Background & Summary

Seagrasses, the unique group of marine angiosperms, belong to the Monocotyledonous order Alismatales and comprise 5 families with 73 species.[1,2]. These flowering grasses inhabit all continents except Antarctica[2], covering 600,000 to 1,600,000 km² of the ocean[3,4,5]. Since they are adapted to submerged survival in highly saline environments and have experienced drastic habitat

transitions from terrestrial to marine, seagrasses are ideal species for studying adaptive evolution[7,10].

Following the *Zostera marina* genome assembly in 2016[7], whole-genome sequences for six seagrass species, including *Cymodocea nodosa*, *Posidonia oceanica*, *Thalassia testudinum*, *Z. marina*[9,10], *Halophila ovalis*[6] and *Z. muelleri*[8] across five families have been published. *Cymodocea nodosa* grows along the temperate Atlantic coast, *P. oceanica* is endemic to the Mediterranean, *T. testudinum* inhabits the Caribbean, while *Z. marina*[2,10], *H. ovalis*[6] and *Z. muelleri*[8] are widely distributed. While *H. ovalis* and *Z. muelleri* currently have only second-generation draft genomes[6,8], the other four seagrass species have chromosomal-level genome assemblies with high genomic integrity, and their haploid chromosome numbers and genome sizes vary from 6 to 18, and 260.5 to 4,261.9 Mb, respectively[9,10].

The assemblies of these genomes have given insights into the comprehension of seagrass evolutionary adaptation by indicating that some species underwent the whole-genome triplication in their early evolutionary phase, which temporality corresponds to the Cenomanian-Turonian anoxic event, while some lineages underwent independent whole-genome duplication (WGD)[10,11]. For instance, after it diverged from the *P. oceanica* lineage, *C. nodosa* underwent a WGD[10]. The widespread phenomenon of transposable element (TE) expansion or explosion in diverse seagrass lineages is also likely associated with the Pleistocene glaciation and the genetic variations resulting from WGD events and has propelled the adaptive evolution of seagrass lineages[10,12]. For example, *T. testudinum* experienced a large-scale Long Terminal Repeat (LTR)/Gypsy burst around 200,000 years ago due to the expansion of its TEs[10].

Gene families associated with terrestrial functions, such as stomatal development, volatile terpenoid synthesis, lignin metabolism, UV protection, and the ethylene signalling pathway, demonstrate significant contraction or even complete loss[10]. On the other hand, the retention and expansion of crucial gene families associated with osmoregulation, salinity adaptation, light harvesting, and carbon acquisition, along with alterations in cell wall polysaccharide composition,

jointly form the physiological foundation for their survival in high-salinity marine environments[10]. Notably, certain lineages undergo species-specific gene losses during evolution, like the loss of NADH dehydrogenase complex genes in *H. ovalis*[6]. Therefore, the genome analyses of these six representative seagrass species provide abundant molecular evidence for convergent evolution across lineages and explain habitat adaptation[6,8,10].

Distinct from other seagrasses, *Phyllospadix* is the only species that has adapted to rocky coast habitats and exhibits a unique dioecious trait within the Zosteraceae family (Fig. 1a)[13]. Its morphological features show multiple adaptations to the turbulent environment of rocky shores, including shorter stolon internodes, a greater number of fibrous roots and dense root hairs that enhance its anchorage ability(Fig. 1c)[14], and a higher-proportion endosperm, a harder seed coat, and special anchoring structures, which improve the survival ability of its seeds(Fig. 1b)[15]. However, the elucidation of the divergence process of the *Phyllospadix* within the Zosteraceae family and the more comprehensive exposition of the adaptive evolutionary mechanisms of seagrass groups depend on the availability of its high-quality genome.

In the present study, we built a genome at the chromosome level for *Phyllospadix iwatensis* through integration of 49.87 Gb Illumina short-reads, 16.76 Gb PacBio HiFi data, and 40.83 Gb high-throughput chromosome conformation capture (Hi-C) data, with a size of 340.56 Mb. The *P. iwatensis* genome featured a contig N50 of 30.64 Mb and scaffold N50 of 33.59 Mb, with 96.44% of genomic sequences mapped to 10 chromosomes. (Fig. 2).

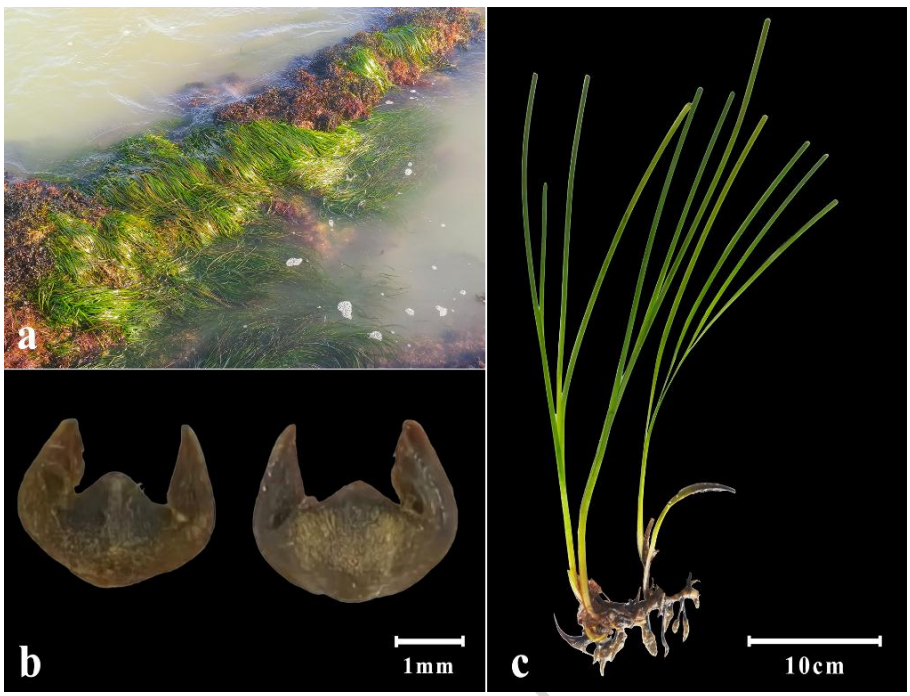

Fig. 1 Habitat and morphology of *P. iwatensis*. a, *P. iwatensis* growing on rocks; b, Morphology of *P. iwatensis* seeds; c, Morphology of the whole plant of *P. iwatensis*.

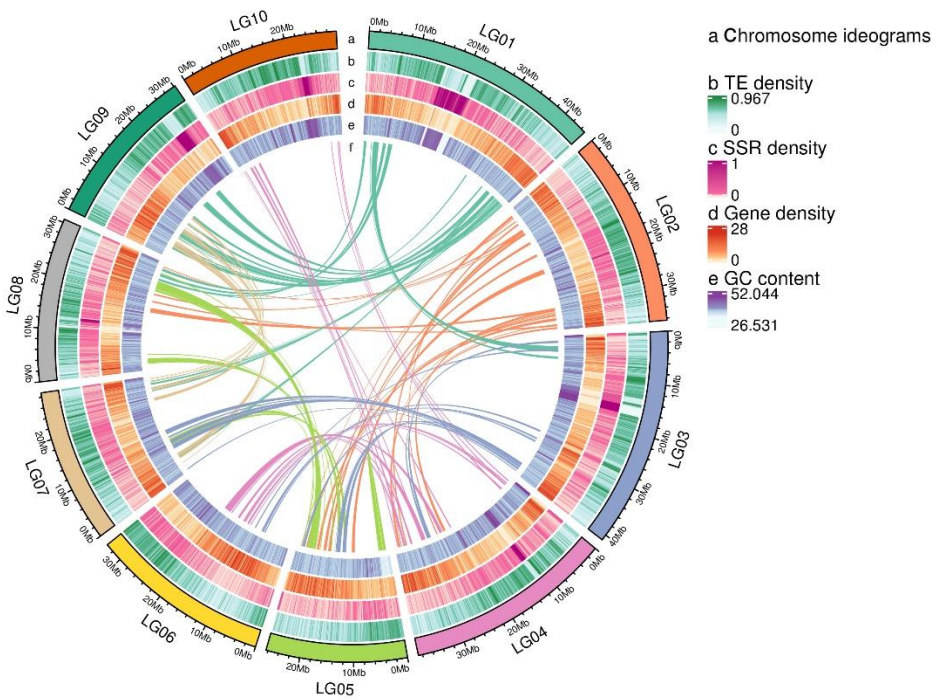

Fig. 2 Circle plot of *P. iwatensis* genome assembly and annotation.

Methods

Sample collection and preparation. An endemic species to Asia, *P. iwatensis* is primarily distributed along the northeastern coast of South Korea, the eastern coastal waters of Japan, and the northern coastal regions of China. Due to its wide distribution in the Yellow and Bohai Sea regions, *P. iwatensis* was selected as the experimental material, while its samples were collected from Nan Changshan Island (120.75795°E, 37.91817°N), Shandong Province, China. Samples were repeatedly cleaned with sterile absorbent cotton wetted in sterilized seawater to ensure their purification. Genomic DNA from samples was extracted via the cetyltrimethylammonium bromide method, whereas RNA isolation employed the TianGen DP411 Kit (DP411, TianGen, China).

Genome survey. Genomic DNA was randomly fragmented using ultrasonic sonication to generate 350 bp fragments, followed by end repair, adenine tailing, adapter ligation, size selection of 350 ± 50 bp, and PCR amplification for short-read library construction. For RNA library preparation, mRNA was isolated with Oligo(dT) beads, fragmented, and used for double-stranded cDNA synthesis, prior to end repair, adapter ligation, and PCR amplification. After being sequenced on the Illumina NovaSeq 6000 platform, 49.87 Gb clean data was generated, aligning with the coverage of 141.60×. Short-read data generated by the Illumina platform underwent quality filtering via fastp.[18]. Filtered high-quality reads served to gauge genome size, with 21-kmers tallied via Jellyfish.[16] software and calculated the genome characteristics using Genomescope[17] software. The genome size of *P.iwatensis* was estimated to be 352.20 Mb, with a heterozygosity rate of approximately 0.14% and a repetitive sequence content of approximately 43.44% (Fig. 3). Additionally, the flow cytometry analysis using the *Solanum lycopersicum* genome[49,50] as a reference revealed that the genome size of *P. iwatensis* was approximately 360 Mb, which was consistent with the k-mer assessment.

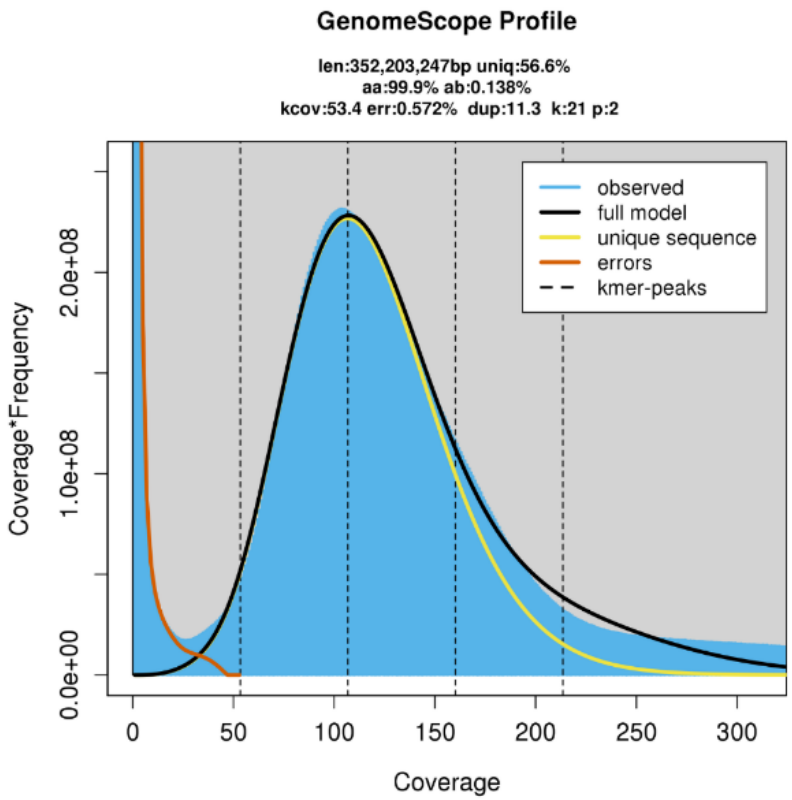

Fig. 3 The genome size of *P. iwatensis* was estimated via 21-mer using raw Illumina short-reads.

Genome Assembly. To construct single-molecule real -time PacBio libraries, the SMRTbell® Express Template Prep Kit 2.0 (PacBio, CA, USA) was employed. These libraries were subsequently sequenced in circular consensus sequencing (CCS) mode using the PacBio Sequel II platform. This process yielded 16.76 Gb of clean data, corresponding to roughly 47x genome coverage for *P. iwatensis*. The chromosomal-level genome of *P. iwatensis* was determined to be 353.13 Mb in size, with a Contig N50 of 30.64 Mb. Additionally, library construction and sequencing were performed using the in situ Hi-C protocol[21].

Genomic mapping of Hi-C data was carried out with Juicer, followed by clustering with 3d-dna[19,20]（Fig. 4）. After assembly and manual heatmap adjustment, 340.56 Mb genomic sequence was anchored to 10 chromosomes, accounting for 96.44% of the genome（Table 1）.

Group	Cluster Num	Cluster Len

LG01	1	44,888,761
LG02	1	36,397,520
LG03	1	41,080,448
LG04	1	38,780,662
LG05	7	26,360,939
LG06	2	33,592,404
LG07	1	28,810,059
LG08	1	30,637,857
LG09	3	30,269,056
LG10	1	29,746,127
Total (Ratio %)	19 (11.59)	340,563,833 (96.44)

Table 1 The assembly of *P. iwatensis* resulted in 10 pseudochromosomes with lengths.

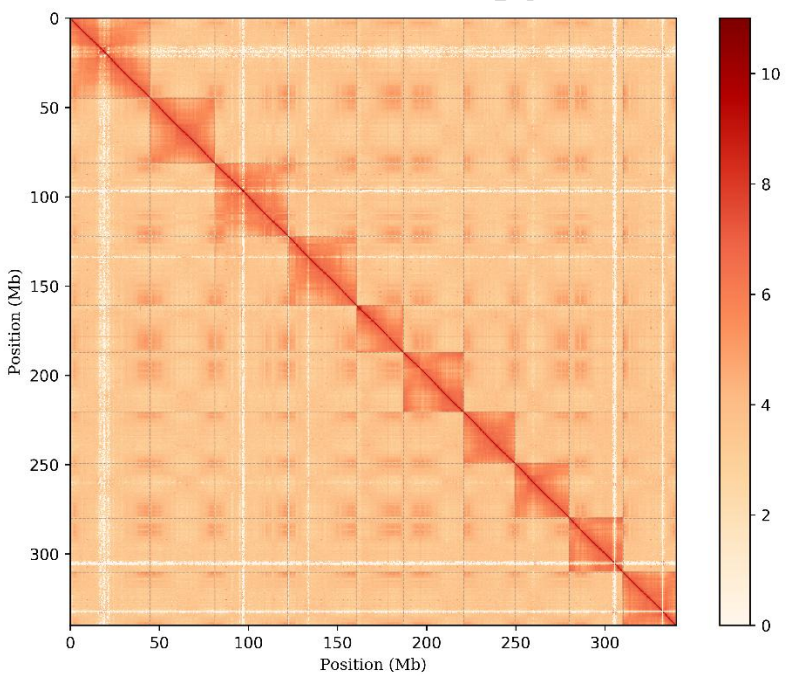

Fig. 4 Hi-C interaction heat map of *P. iwatensis* genome. The color gradient from light to dark corresponds to increasing interaction intensity.

Repeat annotation. The *de novo* prediction was performed using RepeatModeler2 v2.0.1[22], which employed two *de novo* prediction software tools, RECON v1.0.8[23] and RepeatScout v1.0.6[24], and used RepeatClassifier with the Dfam v3.5 known database to classify the prediction results. The LTRs were predicted de novo using LTR_retriever 2.9.0[25], which leverages predictions from

LTRharvest v1.5.10[26] and LTR_FINDER v1.07[27]. The de novo prediction findings were next merged with existing databases; following the removal of duplicates, a repeat sequence database specific to the species was created. RepeatMasker v4.1.2[28] was employed to identify TE sequences, with predictions drawing on the built repeat sequence database. There are 332,323 repetitive sequences were detected, spanning 186,829,029 bp and representing 52.91% of the genome (Table 2). LTR elements were the most prevalent category, covering 167,127 bp and making up 31.06% of the genome. Tandem repeats were predicted via Microsatellite Identification Tool MISA v2.1[29] and Tandem Repeat Finder v409[30]. Around 34,927,161 bp of tandem repeats were ultimately obtained, constituting 9.89% of the genome. (Table 3).

Type	Number	Length (bp)	Percentage (%)
SINE	6,690	956,771	0.27
LINE	27,082	11,589,564	3.28
LTR	167127	109676455	31.06
DNA	131,417	64,605,720	18.30
Unknown	7	519	0.00
Total	332,323	186,829,029	52.91

Table 2 Statistics of the repetitive sequences in *P. iwatensis* genome.

Type	Number	Length	Rate (%)
microsatellite (1-9 bp units)	66,563	1,289,688	0.37
minisatellite (10-99 bp units)	115,085	20,015,743	5.67
satellite (>=100 bp units)	16,219	13,621,730	3.86
Total	197,867	34,927,161	9.89

Table 3 Statistics of tandem repetitive sequences in *P. iwatensis* genome.

Prediction of Genes and Their Functional Annotation. Protein-coding genes were predicted from the repeat-masked genome using a combination of ab initio, homology-based, and transcript-assisted prediction. The *ab initio* prediction was conducted using Augustus v3.1.0[31] and SNAP 2006-07-28[32], while homology-based alignment was done using GeMoMa v1.7[33] with reference gene models from *Arabidopsis thaliana*[47], *Spirodela polyrhiza*[48], *Z. marina*[9], *Z. muelleri*[8], *C. nodosa*, *P. oceanica* and *T. testudinum*[10]. For transcript-based gene annotation, RNA sequencing was performed on six distinct tissues of *P. iwatensis*, specifically leaves, stolons, fibrous roots,

female inflorescences, male inflorescences, and fertilized female inflorescences. To predict the second-generation transcriptome, transcripts were mainly assembled and predicted in two ways. One way involved obtaining transcripts using Hisat v2.1.0[34] and Stringtie v2.1.4[35], and then using GeneMarkS-T v5.1[36] for gene prediction, while the other way included assembling transcripts through RNA-Bloom v2.0.0[37], and then using PASA v2.4.1[38] for gene prediction. For the third-generation transcriptome, gmap (2020-06-30) was used for alignment. After a series of splice site processing, we employed PASA v2.4.1[38] for gene prediction. Prediction outcomes from the three aforementioned methods were integrated via EVM v1.1.1[39], with refinement done using PASA v2.4.1[38] (Fig. 5). Subsequently, the functions of possible genes were annotated by identifying the best matches to proteins in the GO, KEGG[41], KOG, Pfam[43], SWISS-PROT[42], TrEMBL[42], eggNOG[40], and NR databases. In total, 23,198 protein-coding genes were identified in the *P. iwatensis* genome, with each gene having an average length of 4145.94 bp. Among these, 21,954 (94.64%) were matched with functional information (Table 4).

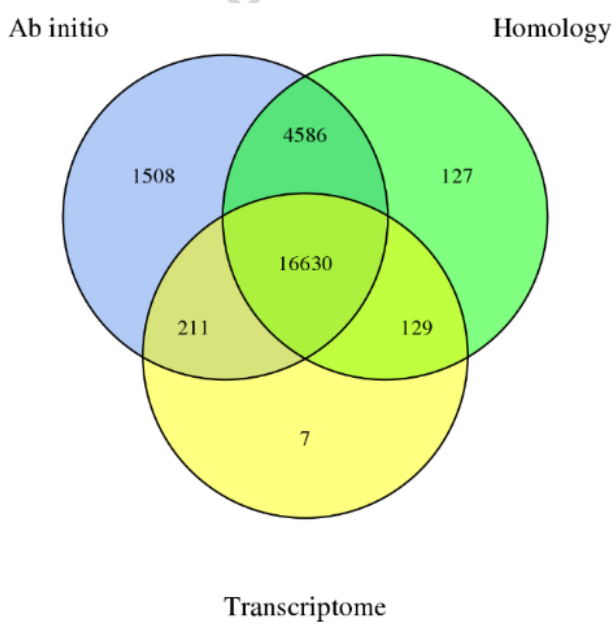

Fig. 5 Distribution map of integrated genes obtained from three prediction approaches

	Database	Total gene number	Percentage (%)
Total		23,198	
Annotated	GO	18,031	77.73
	KEGG	16,236	69.99
	KOG	12,030	51.86
	Pfam	19,181	82.68
	Swissprot	15,780	68.02
	TrEMBL	21,668	93.40
	eggNOG	18,564	80.02
	NR	20,993	90.49
Unannotated		1,244	5.36
Total annotated		21,954	94.64

Table 4 Statistics of functional annotation of *P. iwatensis* genome

Data Records

The complete dataset of *P. iwatensis*, including raw sequencing data and the assembled genome, is available from the National Center for Biotechnology Information (NCBI) and figshare. Raw data encompassing Illumina, PacBio, Hi-C, and RNA sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under the accession numbers SRR34629676[51], SRR34629675[52], SRR34629674[53], and SRR34629673[54]. The genome assembly has been submitted to the NCBI GenBank database with the accession number JBTXFO000000000.1[55]. In addition, the chromosome-level genome assembly of *P. iwatensis*, along with its predicted coding sequences and protein sequences, can be accessed via figshare[56].

Technical Validation

Assessment of sample contamination. High-quality genomes rely on sample purity. To assess contamination of the extracted sample DNA, 10,000 single-end reads were randomly selected from the 350 bp library generated by sequencing, followed by BLAST[44] alignment against the Non-redundant nucleotide (NT) database. Using ncbi-blast+ version 2.2.29 with default parameters.

The alignment results indicated that the samples were free of contamination.

Assessment of genome assembly. Genome assembly quality was evaluated using multiple metrics via the Core Eukaryotic Genes Mapping Approach[45] (CEGMA), which employs a core gene library containing 458 conserved genes from eukaryotic model organisms. Combined with tblastn, genewise, and geneid, CEGMA evaluated the assembled genome, yielding an integrity score of 98.25%. Subsequently, Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.2.1[46] was applied to select the Eukaryota database from the orthologous database OrthoDB 10 (http://cegg.unige.ch/orthodb), indicating an assessed integrity of 98.82%. BUSCO v5.2.2 was employed to evaluate the completeness of gene prediction results. Interestingly, 99.6% of BUSCO genes were identified, indicating that gene prediction is highly complete.

Code Availability

All bioinformatics analyses in this study were performed in strict accordance with the guidelines of the respective tools. No custom scripts were developed; all operations adhered to the standard protocols of the employed software. These tools are publicly accessible, with detailed information on their versions and parameter settings provided in the Methods section.

Data Availability

The complete dataset of *P. iwatensis*, including raw sequencing data (Illumina, PacBio, Hi-C, and RNA sequencing reads) and the assembled genome, is publicly available via the following repositories:

NCBI SRA:

https://identifiers.org/ncbi/insdc.sra:SRR34629676

https://identifiers.org/ncbi/insdc.sra:SRR34629675

https://identifiers.org/ncbi/insdc.sra:SRR34629674

https://identifiers.org/ncbi/insdc.sra:SRR34629673

NCBI GenBank: https://identifiers.org/ncbi/insdc:JBTXFO000000000.1

Figshare: https://doi.org/10.6084/m9.figshare.29649677

References

1. Short, F. T. *et al.* Extinction risk assessment of the world's seagrass species. *Biol. Conserv.* **144**, 1961–1971 (2011).
2. Unsworth, R. K. F., Cullen-Unsworth, L. C., Jones, B. L. H. & Lilley, R. J. The planetary role of seagrass conservation. Science. **377**, 609–613 (2022).
3. McKenzie, L. J. *et al.* The global distribution of seagrass meadows. *Environ. Res. Lett.* **15**, 74041 (2020).
4. Duffy, J. E. *et al.* Toward a Coordinated Global Observing System for Seagrasses and Marine Macroalgae. *Front. Mar. Sci.* **6**, 317 (2019).
5. Gallagher, A. J. *et al.* Tiger sharks support the characterization of the world's largest seagrass ecosystem. *Nat. Commun.* **13**, 6328 (2022).
6. Lee, H. *et al.* Genomic comparison of two independent seagrass lineages reveals habitat-driven convergent evolution. *J. Exp. Bot.* **69**, 3689–3702 (2018).
7. Olsen, J. L. *et al.* The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. *Nature* **530**, 331–335 (2016).
8. Lee, H. *et al.* The genome of a Southern Hemisphere seagrass species (Zostera muelleri). *Plant Physiol.* **172**, 272–283 (2016).
9. Ma, X. *et al.* Improved chromosome-level genome assembly and annotation of the seagrass, Zostera marina (eelgrass). *F1000Research* **10**, 289 (2021).
10. Ma, X. *et al.* Seagrass genomes reveal ancient polyploidy and adaptations to the marine environment. *Nat. Plants* **10**, 240–255 (2024).
11. Van De Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).
12. Böse, M., Lüthgens, C., Lee, J. R. & Rose, J. Quaternary glaciations of northern Europe. *Quat. Sci. Rev.* **44**, 1–25 (2012).
13. Sullivan, B. K. & Short, F. T. Taxonomic revisions in Zosteraceae (Zostera, Nanozostera, Heterozostera and Phyllospadix). *Aquat. Bot.* **187**, 103636 (2023).
14. Cao, Z., Zhang, W. & Zhao, H. Morphology and anatomy of *Phyllospadix iwatensis* Makino and their adaptation to marine environment. *Oceanol. Limnol. Sin.* **46**, 1326–1332 (2015).
15. Han, T. *et al.* The seed morphology and internal characteristics of seagrass, surfgrass

Phyllospadix iwatensis. *J. Trop. Oceanogr.* **41**, 105–113 (2022).

16. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
17. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).
18. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
19. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
20. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
21. Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
22. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* **117**, 9451–9457 (2020).
23. Bao, Z. & Eddy, S. R. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* **12**, 1269–1276 (2002).
24. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
25. Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).
26. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008).
27. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
28. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma.* **4**, 4.10.1-4.10.14 (2009).
29. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: a web server for microsatellite prediction. *Bioinformatics* **33**, 2583–2585 (2017).
30. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
31. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
32. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).

33. Keilwagen, J. *et al.* Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89–e89 (2016).
34. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
35. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
36. Tang, S., Lomsadze, A. & Borodovsky, M. Identification of protein-coding regions in RNA transcripts. *Nucleic Acids Res.* **43**, e78–e78 (2015).
37. Nip, K. M. *et al.* RNA-Bloom enables reference-free and reference-guided sequence assembly for single-cell transcriptomes. *Genome Res.* **30**, 1191–1200 (2020).
38. Haas, B. J. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
39. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
40. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
41. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).
42. Boeckmann, B. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370 (2003).
43. Finn, R. D. Pfam: clans, web tools and services. *Nucleic Acids Res.* **34**, D247–D251 (2006).
44. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
45. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
46. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
47. Hou, X., Wang, D., Cheng, Z., Wang, Y. & Jiao, Y. A near-complete assembly of an Arabidopsis thaliana genome. *Mol. Plant* **15**, 1247–1250 (2022).
48. An, D. *et al.* Plant evolution and environmental adaptation unveiled by long-read whole-genome sequencing of Spirodela. *Proc. Natl. Acad. Sci.* **116**, 18893–18899 (2019).
49. The Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy

fruit evolution. *Nature* **485**, 635–641 (2012).
50. Su, X. *et al.* A high-continuity and annotated tomato reference genome. *BMC Genomics* **22**, 898 (2021).
51. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR34629676 (2026).
52. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR34629675 (2026).
53. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR34629674 (2026).
54. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR34629673 (2026).
55. *NCBI GenBank* https://identifiers.org/ncbi/insdc: JBTXFO000000000.1 (2026).
56. Wang, J. The chromosomal-level genome assembly and annotation of *Phyllospadix iwatensis* (Surfgrass). *Figshare* https://doi.org/10.6084/m9.figshare.29649677 (2026).

Acknowledgements

This reaserch was supported by the National Natural Science Foundation of China (NO. 42476112) and the Shandong Provincial Bureau of Geology and Mineral Resources project (NO. HJ202510).

Author Contributions

L.Z.N. and Z.Q.S. conceived and designed the study, secured funding, and participated in manuscript writing, review, and editing. W.D.W. and Z.K. conducted the experiments and analyzed the data. W.J.Y. analyzed the data and drafted the initial manuscript. All authors reviewed and approved the final version of the manuscript.

Competing Interests

The authors declare no competing interests.

Group	Cluster Num	Cluster Len
LG01	1	44,888,761
LG02	1	36,397,520
LG03	1	41,080,448
LG04	1	38,780,662

LG05	7	26,360,939
LG06	2	33,592,404
LG07	1	28,810,059
LG08	1	30,637,857
LG09	3	30,269,056
LG10	1	29,746,127
Total (Ratio %)	19 (11.59)	340,563,833 (96.44)

Table 1 The assembly of *P. iwatensis* resulted in 10 pseudochromosomes with lengths

Type	Number	Length (bp)	Percentage (%)
SINE	6,690	956,771	0.27
LINE	27,082	11,589,564	3.28
LTR	167127	109676455	31.06
DNA	131,417	64,605,720	18.30
Unknown	7	519	0.00
Total	332,323	186,829,029	52.91

Table 2 Statistics of the repetitive sequences in *P. iwatensis* genome.

Type	Number	Length	Rate (%)
microsatellite (1-9 bp units)	66,563	1,289,688	0.37
minisatellite (10-99 bp units)	115,085	20,015,743	5.67
satellite (>=100 bp units)	16,219	13,621,730	3.86
Total	197,867	34,927,161	9.89

Table 3 Statistics of tandem repetitive sequences in *P. iwatensis* genome.

	Database	Total gene number	Percentage (%)
Total		23,198	
Annotated	GO	18,031	77.73
	KEGG	16,236	69.99
	KOG	12,030	51.86
	Pfam	19,181	82.68
	Swissprot	15,780	68.02

	TrEMBL	21,668	93.40
	eggNOG	18,564	80.02
	NR	20,993	90.49
Unannotated		1,244	5.36
Total annotated		21,954	94.64

Table 4 Statistics of functional annotation of *P. iwatensis* genome