



OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly for the ecologically and economically important alga *Saccharina japonica*

Xiaodong Li^{1,2} , Yu-Long Li^{3,4}, Chenhui Zhong⁵, Jing Li^{1,2}, Li Su^{1,2}, Jin-Xian Liu^{3,4} & Shaojun Pang^{1,2}

Saccharina japonica is a major kelp species of brown algae with the highest production among aquaculture seaweeds and holds important ecological and economic value. Despite advancements in domestication, a high-quality chromosome-level genome assembly is needed to assist its genetic improvement. Previous genome assemblies of *S. japonica* were either on a draft-level or highly fragmented. Here, we generated a high-quality chromosome-level genome for the female sporophyte using PacBio sequencing and Hi-C. The genome is 516.11 Mb, with contig N50 length of 491.30 Kb and scaffold N50 length of 16.24 Mb, anchored into 32 pseudo-chromosomes. Repetitive elements constituted 45.07% of the genome, and 17,739 protein-coding genes were predicted, of which 82% were functionally annotated. This genome provides a crucial resource for biotechnological advances in *S. japonica* breeding and offers insights into the ecology and evolution of brown algae.

Background & Summary

The brown macroalga *Saccharina japonica* is a major component of intertidal seaweed flora in the north-west coast of the Pacific Ocean, including Russia, Japan, Korea and China¹. It is one of the most economically important seaweeds in aquaculture, being extensively utilized as human food, marine animal feeds and raw industrial materials for various end products². Through artificial cultivation, the output of *S. japonica* reached 10.86 million tons in fresh weight and had a market value of more than 4.4 billion US dollars in 2020³. Over the past decades, this species has become one of the most cultivated macroalgae in China⁴. Using traditional selective breeding and/or hybridization techniques, more than a dozen certified varieties have been successfully bred and farmed in China, which have contributed to the farming success of this cold-water species in both the temperate and subtropical regions along the coasts of China.

High quality genomic information can provide important resources for biotechnological approaches to crop genetic improvements. Despite recent advances in molecular-aided genetic breeding, genetic structure assessments and transcriptomic studies in this important marine farmed “crop”^{4–10}, *S. japonica* remains to be a species with limited genetic engineering tools due to the absence of a high-quality genome assembly. By using Illumina reads and 10 Gb of PacBio long reads, Ye *et al.*¹¹ reported a draft genome of *S. japonica* with assembly size of 537 Mb and contig N50 length of 58.87 Kb. By using the High-throughput Chromosome Conformation Capture (Hi-C) data and Illumina reads, Liu *et al.*¹² assembled a chromosome-level genome of *S. japonica* with genome size of 580.5 Mb and contig N50 length of 4.73 Kb. However, the genome assembly is fragmented and annotations reported in Liu *et al.*¹² are not publicly accessible. Considering the ecological and economic importance of *S. japonica*, a high-quality chromosome level genome assembly is still urgently needed.

¹Key Laboratory of Breeding Biotechnology and Sustainable Aquaculture (CAS), Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China. ²Laboratory for Marine Biology and Biotechnology, Qingdao Marine Science and Technology Center, Qingdao, China. ³CAS Key Laboratory of Marine Ecology and Environmental Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China. ⁴Laboratory for Marine Ecology and Environmental Science, Qingdao Marine Science and Technology Center, Qingdao, China. ⁵Provincial Key Laboratory of Cultivation and High-value Utilization of Marine Organisms, Fisheries Research Institute of Fujian Province, Xiamen, China. ✉e-mail: sjpang@qdio.ac.cn

Items	Statistical result
Sequencing	
Short reads sequencing	
Raw data (Gb)	58.14
Sequencing depth (X)	104
PacBio sequencing	
Raw data (Gb)	39.55
Sequencing depth (X)	71
Hi-C sequencing	
Raw data (Gb)	53.19
Sequencing depth (X)	95
Assembly features	
PacBio sequencing assembly	
Genome size (Mb)	516.11
Number of contigs	2802
Contig N50 (Kb)	491.30
Hi-C assembly	
Genome size (Mb)	516.35
Chromosome number	32
Anchored rate (%)	96.15
Number of scaffolds	405
Scaffold N50 (Mb)	16.24
Genome annotation	
Number of protein-coding genes	17739
Number of genes annotated	14510
Average gene length (bp)	12489
Average exon length (bp)	231
Average exons per gene	6.99
Number of exons	123996
Average intron length (bp)	1816
Average introns per gene	5.99
Number of introns	106257
Total size of TEs (Mb)	215.21
TEs in genome (%)	41.76
Total size of repeated sequences (Mb)	232.3
Repeated sequences (%)	45.07

Table 1. Statistics of *S. japonica* genome assembly and annotations in this study.

In this study, we generated a chromosome-level genome assembly (516.11 Mb) of *S. japonica* female sporophyte using Illumina, PacBio HiFi, and Hi-C data. The assembly exhibited a contig N50 length of 491.30 Kb and scaffold N50 length of 16.24 Mb, indicating superior continuity compared to the previously reported ones. Approximately ~96.15% of the genome was anchored to 32 chromosomes, consistent with the chromosome number reported for species previously classified in the *Laminaria* genus, including *Saccharina japonica*¹³. Repetitive elements constituted 45.07% of the genome assembly. The evaluation indexes of the new assembly such as the Merqury QV (34.86), the completeness (96.67%) and the mapping ratio of pair-end short reads (87.16%) indicate a higher level of genome completeness. A total of 17,739 protein-coding genes were predicted and functional annotations were assigned to 14,510 (82%) of them, much enhanced the interpretative power.

This high-quality chromosome-level *S. japonica* genome, overcoming the limitations of the previous versions, provides a foundational genomic resource for developing molecular approaches in future breeding efforts and offers insights for ecological and evolutionary studies of the brown algae in general.

Methods

Preparation of biological materials. *Gametophyte source and gametogenesis condition.* The female gametophyte strain HJS-C used in this study was zoospore-derived clone preserved in Seaweed Culture Collection Center, Chinese Academy of Sciences (<http://www.caslivealgae.com>). The strain was stored in liquid culture medium in a glass tube in a temperature-controlled incubator at 10 °C under 10 μmol photons m⁻² s⁻¹ white light with a 12 h: 12 h light-dark cycles.

For gametogenesis, the filamentous gametophytes were broken down into 5–10 cell fragments using an electric blender (20,000 r/min. JYL-C012, Jiuyang, China). The fragments were inoculated into 9 cm Petri dishes at an average density of 10–15 fragments per microscopic field (100 × magnification). Gametogenesis was carried

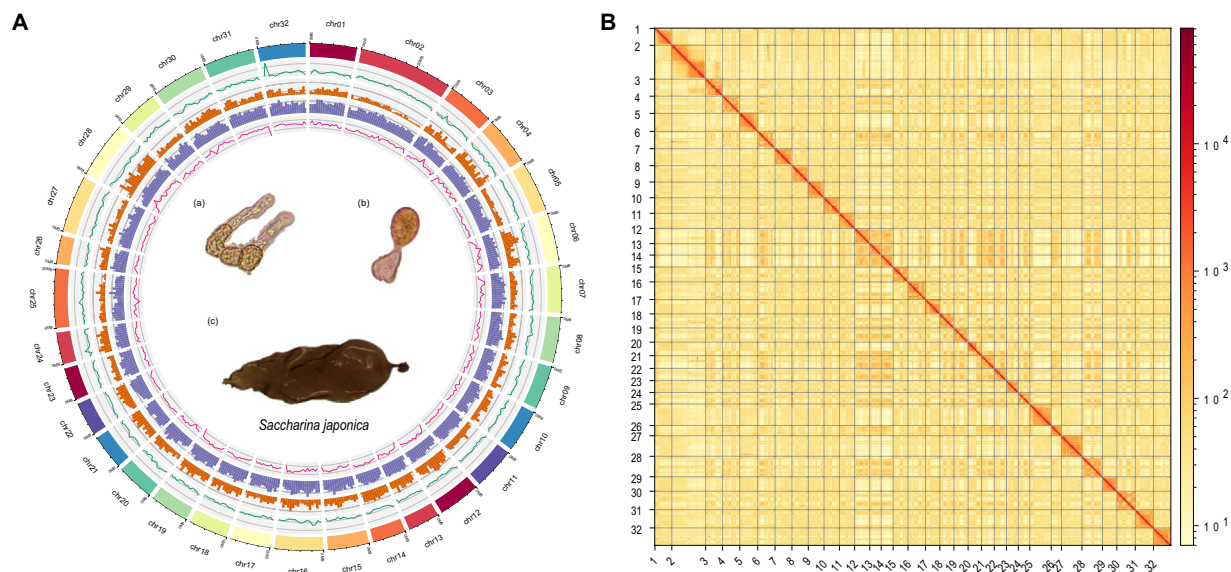


Fig. 1 Chromosome-level genome assembly of *S. japonica*. **(A)** Circos plot of the *S. japonica* genome assembly. From outer to inner layers were chromosomes, depth of short-reads, gene densities, repetitive elements densities and GC contents, respectively. (a,b) female gametophyte, (c) female sporophyte. **(B)** Hi-C contact map for the genome assembly of *S. japonica*.

out at 12 °C, 50 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ in white light in PES¹⁴ with a medium renewal period of three days and a photoperiod of 12:12 (L: D).

Culture of the female sporophyte. When parthenogenetic sporophytes, formed through haploid doubling, appeared in the HJS-C culture petri dishes after 20 day's culture, they were subsequently transferred into a 1-l glass beaker containing PES medium for further culture in aeration using compressed air. The culture beaker was placed in a temperature and light-controlled incubator at 12 °C in 60 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ white light with a photoperiod of 12:12 (L: D). As the sporophytes reached 3–5 cm in length, they were transferred into an 80-l flow-through tank for grown-up. The tank was set in a room with reduced ambient solar irradiance. Daily PAR went up to maximally 300 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ for ca 2 h on a sunny day. The sporophytes were tumbled in the water currents as described in Pang *et al.*¹⁵. Daily water temperature varied between 12 to 14 °C. Every two days, KNO_3 and KH_2PO_4 were added to reach levels of 10 and 1 mg L^{-1} , respectively.

The algal material used for genome sequencing has been preserved in the Seaweed Culture Collection Center, Chinese Academy of Sciences (<http://www.caslivealgae.com>) and is accessible for research purposes according to standard repository procedures.

Sample preparation and genomic DNA extraction. Female sporophytes were harvested when they reached 10 cm in length. To ensure the least amount of contamination during PacBio sequencing, the sporophytes were repeatedly treated with antibiotics until no bacterial colonies would form when plated. For antibiotic treatments, a stock solution of gentamicin was prepared at a final concentration of 200 $\mu\text{g mL}^{-1}$. The pre-washed sporophytes were immersed in this solution for a duration of 2 hours. Following this treatment, the sporophytes were vigorously shaken in autoclaved seawater to dislodge surface bacteria that had been inhibited or killed. Bacteria were determined by plate coating method. For sampling, sporophytes were rapidly frozen with liquid nitrogen, and subsequently preserved at -80°C . Genomic DNA was extracted from the female sporophyte using improved CTAB method.

Library preparation and sequencing. In brief, SMRTbell libraries were constructed for the long read sequencing based on PacBio's standard protocol (Pacific Biosciences, Menlo Park, CA, USA). These libraries were sequenced through a PacBio Sequel II System (Pacific Biosciences). About 2.36 million HiFi reads (39.55 Gb) with a mean length of 16,763 Kb were generated. For the Hi-C sequencing, genomic DNAs were fixed with formaldehyde, sheared by the restriction enzyme MboI to build a Hi-C library, and then sequenced on a NovaSeq6000 sequencing platform (Illumina Inc., San Diego, CA, USA). A total of 53.19 Gb of 150 bp paired-end Hi-C data were generated (Table 1).

Genome assembly. To assemble the genome, hifiasm (v0.19.9-r616)¹⁶ was employed to generate contigs utilizing PacBio HiFi reads using parameters “-hg-size 550 m -l 0”. Contigs sequences were checked for adaptors and contaminants using the NCBI Foreign Contamination Screen (FCS) tools (<https://github.com/ncbi/fcs>) and BlobTools¹⁷. After removing potential adaptors and contaminations, contigs were further polished using

Chromosome ID	No. contigs	Length (bp)	Percentage (%)
1	62	16,237,944	3.14
2	43	32,514,357	6.30
3	80	16,341,770	3.16
4	107	16,504,999	3.20
5	51	17,304,895	3.35
6	107	16,664,931	3.23
7	65	16,247,666	3.15
8	70	15,757,540	3.05
9	61	15,140,275	2.93
10	67	14,870,419	2.88
11	72	14,602,799	2.83
12	106	14,590,997	2.83
13	96	11,013,574	2.13
14	93	11,528,183	2.23
15	77	14,259,657	2.76
16	86	16,910,554	3.28
17	58	13,768,380	2.67
18	85	13,553,254	2.62
19	95	13,627,293	2.64
20	59	12,995,717	2.52
21	87	12,205,725	2.36
22	81	11,781,832	2.28
23	52	11,346,182	2.20
24	75	11,081,855	2.15
25	53	20,615,488	3.99
26	62	9,896,441	1.92
27	49	19,730,202	3.82
28	134	19,736,508	3.82
29	48	13,869,707	2.69
30	87	17,721,060	3.43
31	76	17,475,613	3.38
32	62	16,569,579	3.21
Total	2406	496,465,396	96.15

Table 2. Statistics of chromosome length in *S. japonica*.

high-quality short reads with NextPolish²¹⁸. Subsequently, the assembled contigs were organized into chromosomes utilizing the allhic v0.9.8¹⁹, followed by manual curation with JuiceBox (v2.17)²⁰. To further improve the continuity, the scaffolds were gap-filled with HiFi reads using TGS-GapCloser 1.2.1²¹. The final assembly comprised 32 chromosomes (Fig. 1), collectively spanning 496.47 Mb, accounting for approximately 96.15% of the entire genome assembly, while the individual chromosome lengths ranged from 32.51 Mb (Chr2) to 9.90 Mb (Chr26) (Table 2). Comparing with the previous genome assembly of *S. japonica* presented by Ye *et al.*¹¹ and Liu *et al.*¹², our assembly has much longer contig and scaffold N50 (Table 3), which indicated the high-quality of our assembly, with superior continuity.

Genome annotation. Genes were predicted using evidence from both ab initio gene predictors and protein and transcript alignments. First, a de novo repeat library was constructed using EDTA v2.2.1²², and repeats were identified using RepeatMasker v4.1.2 (<https://repeatmasker.org/RepeatMasker/>). Repeat sequences were soft masked before further gene predictions. Secondly, Braker pipeline v3.0.3²³ was used to train the gene prediction tools GeneMark-ETP²⁴ and Augustus v3.5.0²⁵ and to generate the ab initio predictions based on RNA-seq data and proteins from OrthoDB v11²⁶. Thirdly, RNA-seq reads (The raw data is deposited in NCBI, with Bioproject accession of PRJNA1181304) were assembled using Trinity v2.15.1²⁷, and the results were passed to PASA pipeline v2.5.3²⁸ to generate high quality gene structures. Fourth, evidence from protein sequences of OrthoDB v11 and RNA-seq reads was extracted from the alignments to the genome by using MetaEuk²⁹, Minimap2 v2.26-r1175³⁰ and StringTie v2.2.1³¹. Finally, the above ab initio gene predictions and protein and transcript alignments were combined into weighted consensus gene structures using EVIDENCEModeler v2.1.0³² implemented by Funannotate v1.8.17³³ (<https://github.com/nextgenusfs/funannotate>). Protein functions were annotated using Funannotate from alignments on multiple databases including the Swiss-Prot/TrEMBL³⁴, Pfam-A³⁵, EggNOG³⁶ and InterProScan³⁷. A total of 17,739 protein-coding genes are predicted, of which 14,510 (82%) genes are functionally annotated. The statistics of gene models, including gene length, CDS length, intron length and intron length in *S. japonica* were comparable to other Laminariales algae (Fig. 2).

Statistics	<i>S. japonica</i> (This study)	<i>S. japonica</i> (Liu <i>et al.</i> ¹²)	<i>S. japonica</i> (Ye <i>et al.</i> ¹¹)
Platform	Illumina + PacBio + Hi-C	Illumina + Hi-C	Illumina + PacBio
Genome size (Mb)	516.11	580.5	537.64
Raw data of short reads sequencing (Gb)	58.14	76.46	84
Raw data of PacBio sequencing (Gb)	39.55	/	10
Contig N50 (Kb)	491.30	4.73	58.87
Scaffold N50 (Mb)	16.24	13.64	0.25
Number of contigs	2802	418683	29670
Number of scaffolds	405	236802	13327
Hi-C anchored rate(%)	96.15	89.19	/
Number of protein-coding genes	17739	35725	18733
Average gene length (bp)	12489	5591	9587
Average exon per gene	6.99	4.68	6.54
Repeated sequences (%)	45.07	46.03	39

Table 3. Statistics of *S. japonica* genome assembly and annotations.

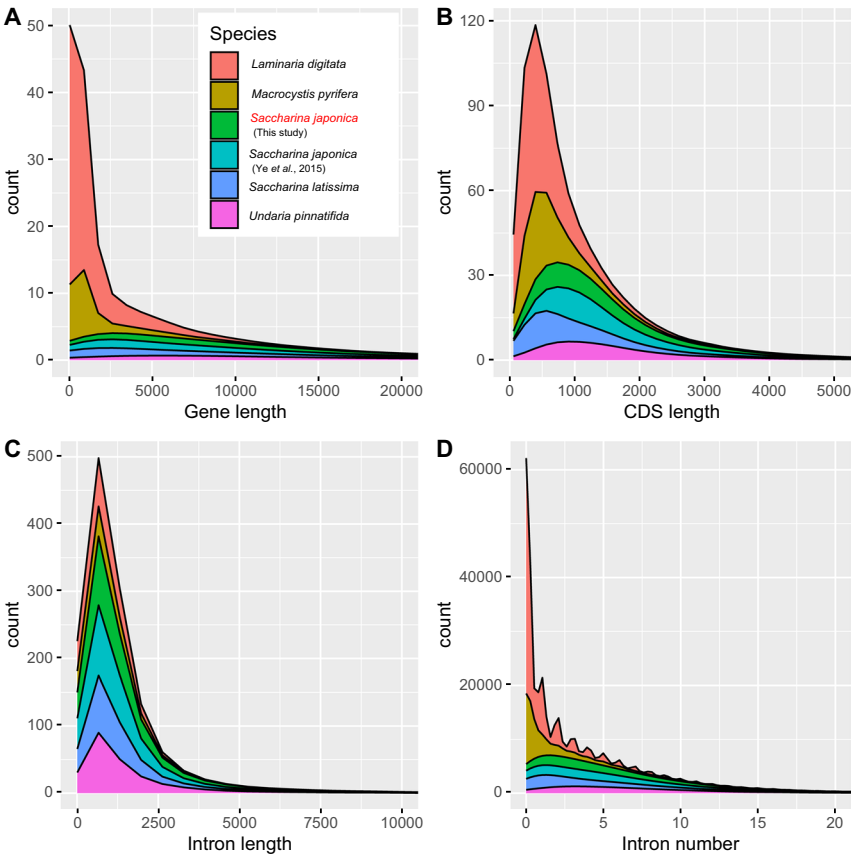


Fig. 2 The composition of gene elements in *S. japonica* and other closely related species. **(A)** Distribution of gene length. **(B)** Distribution of CDS length. **(C)** Distribution of intron length. **(D)** Distribution of intron number.

Data Records

The sequencing data from Illumina has been deposited in Sequence Read Archive (SRA) database at NCBI (SRX26590917)³⁸.

The sequencing data from PacBio has been deposited in Sequence Read Archive (SRA) database at NCBI (SRX26590916)³⁹.

The sequencing data from Hi-C has been deposited in Sequence Read Archive (SRA) database at NCBI (SRX26590918)⁴⁰.

The assembly genome data was deposited in GenBank at NCBI (JBIYRG000000000.1)⁴¹.

The genomic assembly and annotation results were deposited in the Figshare database (<https://doi.org/10.6084/m9.figshare.27569472>)⁴².

Technical Validation

The quality of the genome assembly was checked using a k-mer based method as implemented in Merquy v1.3⁴³, resulting a QV score of 34.86 and completeness of 96.67%. We further mapped the paired-end short reads to the genome using bwa-mem2 v2.2.1⁴⁴, resulting a mapping ratio of 87.16% and proper paired mapping ratio of 83.62%.

Code availability

No custom code was used in this study. The data analyses used standard bioinformatic tools specified in the methods.

Received: 12 November 2024; Accepted: 11 February 2025;

Published online: 18 February 2025

References

- Xu, D. *et al.* Effects of CO₂ and Seawater Acidification on the Early Stages of *Saccharina japonica* Development. *Environ. Sci. Technol.* **49**, 3548–3556 (2015).
- Tseng, C. & Wu, C. *The Kelp Cultivation*. (Science Press, 1962).
- FAO. *The State of World Fisheries and Aquaculture 2024*. <https://doi.org/10.4060/cd0683en> (FAO, 2024).
- Li, X. *et al.* Genetic structure, major chemical component contents and agronomic characters of the principal cultivars of the kelp *Saccharina japonica* in the largest kelp production province in China. *J Appl Phycol* **35**, 763–772 (2023).
- Hu, Z. *et al.* Kelp aquaculture in China: a retrospective and future prospects. *Reviews in Aquaculture* **13**, 1324–1351 (2021).
- Hwang, E. K., Yotsukura, N., Pang, S., Su, L. & Shan, T. Seaweed breeding programs and progress in eastern Asian countries. *Phycologia* **58**, 484–495 (2019).
- Li, X. *et al.* *Saccharina* sporophytes settling on kelp cultivation rafts come from wild intertidal populations and not from the farmed cultivars. *Botanica Marina* <https://doi.org/10.1515/bot-2023-0096> (2024).
- Lin, Z. *et al.* Integrated transcriptomic and proteomic analysis unravels the response mechanisms of *Saccharina japonica* to heat stress. *Aquaculture* **595**, 741528 (2025).
- Su, L. *et al.* Aquaculture of the hybrid cultivars of *Saccharina japonica*: Removing the obstacle of sori production by photoperiodic control. *Aquaculture* **519**, 734917 (2020).
- Zhang, Y., Wang, X., Shan, T., Pang, S. & Xu, N. Transcriptome profiling of the meristem tissue of *Saccharina japonica* (Phaeophyceae, Laminariales) under severe stress of copper. *Marine Genomics* **47**, 100671 (2019).
- Ye, N. *et al.* *Saccharina* genomes provide novel insight into kelp biology. *Nature Communications* **6**, 6986 (2015).
- Liu, T. *et al.* Evolution of Complex Thallus Alga: Genome Sequencing of *Saccharina japonica*. *Front. Genet.* **10** (2019).
- Yabu, H. Chromosome number in four species of *Laminaria* (Phaeophyta). *Jpn J. Phycol.* **39**, 185–187 (1991).
- Hill, M. N. *The Sea, Volume 2: The Composition of Sea-Water; Comparative and Descriptive Oceanography*. (Harvard University Press, 1963).
- Pang, S. & Lüning, K. Tank cultivation of the red alga *Palmaria palmata*: Effects of intermittent light on growth rate, yield and growth kinetics. *Journal of Applied Phycology* **16**, 93–99 (2004).
- Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021).
- Challis, R., Richards, E., Rajan, J., Cochrane, G. & Blaxter, M. BlobToolKit – Interactive Quality Assessment of Genome Assemblies. *G3 Genes[Genomes]Genetics* **10**, 1361–1374 (2020).
- Hu, J. *et al.* NextPolish2: A Repeat-aware Polishing Tool for Genomes Assembled Using HiFi Long Reads. *Genomics, Proteomics & Bioinformatics* **22**, qzad009 (2024).
- Zhang, X., Zhang, S., Zhao, Q., Ming, R. & Tang, H. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat. Plants* **5**, 833–845 (2019).
- Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *cells* **3**, 99–101 (2016).
- Xu, M. *et al.* TGS-GapCloser: A fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *GigaScience* **9**, g1aa094 (2020).
- Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* **20**, 275 (2019).
- Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-Genome Annotation with BRAKER. in *Gene Prediction: Methods and Protocols* (ed. Kollmar, M.) 65–95, https://doi.org/10.1007/978-1-4939-9173-0_5 (Springer, New York, NY, 2019).
- Brüna, T., Lomsadze, A. & Borodovsky, M. GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Res.* **34**, 757–768 (2024).
- Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research* **33**, W465–W467 (2005).
- Kuznetsov, D. *et al.* OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research* **51**, D445–D451 (2023).
- Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644–652 (2011).
- Haas, B. J. *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research* **31**, 5654–5666 (2003).
- Levy Karin, E., Mirdita, M. & Söding, J. MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. *Microbiome* **8**, 48 (2020).
- Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
- Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290–295 (2015).
- Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**, R7 (2008).
- Agudelo-Valencia, D., Uribe-Echeverry, P. T. & Betancur-Pérez, J. F. *De novo* assembly and annotation of the *Ganoderma australe* genome. *Genomics* **112**, 930–933 (2020).
- Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL. *Nucleic Acids Research* **25**, 31–36 (1997).
- Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Research* **42**, D222–D230 (2014).
- 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 Research* **47**, D309–D314 (2019).
- Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
- NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRX26590917> (2024).

39. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRX26590916> (2024).
40. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRX26590918> (2024).
41. NCBI GenBank <https://identifiers.org/ncbi/insdc:JBIYRG000000000> (2025).
42. Li, X. Genome assembly and annotation of *Saccharina japonica* female sporophyte. *figshare* <https://doi.org/10.6084/m9.figshare.27569472> (2025).
43. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merquy: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology* **21**, 245 (2020).
44. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. in *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* 314–324, <https://doi.org/10.1109/IPDPS.2019.00041> (2019).

Acknowledgements

This work was supported by the National Key R&D Program of China (2024YFD2402700); the Taishan Industrial Program (ts20241109); a Key Technology Research and Development Program of Shandong Province (2021LZGC029, 2023LZGCQY002, 2021LZGC004); a Key Supporting Project from the Centre for Ocean Mega-Science, Chinese Academy of Sciences (COMS2020J01); China Agriculture Research System (CARS-50); Biological Resources Program, Chinese Academy of Sciences (KFJ-BRP-004); the Open Program of Key Laboratory of Cultivation and High-value Utilization of Marine Organisms in Fujian Province (2023fjscq01); Fujian Provincial Bureau of Ocean and Fisheries (FJHY-YYKJ-2024-2-1). The authors are grateful to Xiaofei Leng and Mingfu Zhang for their kind help in collecting samples.

Author contributions

Xiaodong Li and Shaojun Pang designed the research. Shaojun Pang collected samples. Xiaodong Li and Jing Li cultured the sequencing samples. Xiaodong Li, Yu-Long Li, Chenhui Zhong, Li Su and Jin-Xian Liu performed experiments or data analyses; Xiaodong Li, Yu-Long Li, Jin-Xian Liu and Shaojun Pang wrote and revised the manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025