scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-scale genome assembly and annotation of *Saccharina sculpera*

Lihao Guo[1], Mingyu Zhong[1], Weiwei Wang[2,3], Xiaojie Li[2,4,5] ✉ & Quansheng Zhang[1,5] ✉

Saccharina sculpera **is a brown alga (kelp) that is rich in fucoidan (a sulphated heteropolysaccharide) and has shown great potential for medical applications. However, fucoidan accumulation mechanisms in *S. sculpera* have not yet been elucidated owing to the absence of its genome assembly. In this research, we generated a chromosome-level genome of *S. sculpera* using Illumina, PacBio HiFi, and high-throughput chromosome conformation capture sequencing technologies. The assembled genome was 726.56 Mb in size, with a contig N50 of 1.08 Mb and scaffold N50 of 19.12 Mb. Approximately 98% of the genomic sequences were ultimately anchored to 34 pseudo-chromosomes. Totally, 25,685 protein-coding genes were identified, of which 86.42% were functionally annotated. Additionally, 438.5 Mb repeat sequences were predicted, which accounted for 60.26% of the assembled genome. The chromosome-level genome of *S. sculpera* offers an invaluable genetic resource for investigating the molecular mechanism underlying fucoidan accumulation in brown algae.**

Background & Summary

Saccharina sculpera is a perennial brown alga (kelp) species that belongs to the family Laminariaceae in the order Laminariales of class Phaeophyceae[1]. It is naturally distributed along the coasts of Gangneung (Korea), Hokkaido (Japan), and southern Sakhalin (Russia)[2]. Juvenile *S. sculpera* sporophytes are characterized by thin, olive-coloured, membranous blades, while the mature sporophytes feature elongated, leathery, dark brown blades, which range between 1–2.5 m in length and 0.3–0.4 m in width[3]. The mature *S. sculpera* blades have a unique, uneven cloud-like pattern that spans their entire surface[4]. Remarkably, *S. sculpera* is significantly rich in fucoidan content, which is approximately three times higher than that of the major cultivated species, *Saccharina japonica*[5–7].

Saccharina is one of the most cultivated and commercially significant brown algae species globally[8]. According to the Food and Agriculture Organization of the United Nations, the global aquaculture production of kelp in 2022 was approximately 10.86 million tonnes, with China accounting for 89% of the production, valued at US$3.8 billion[9]. Kelp is primarily utilized in the food, chemical, and medicine industries. More specifically, kelp is an essential component of Asian cuisine, particularly in Korea and Japan, owing to its high nutritional value and distinctive flavour profile. It is consumed directly, incorporated in prepared foods, or used as a food additive in the culinary field[10,11]. Kelp also serves as a raw material for the industrial production of alginate and mannitol, which function as gelling, stabilizing, and thickening agents in multiple industrial sectors, including cosmetics, textiles, paper, dairy products, and paints[12,13]. Moreover, other bioactive compounds extracted from brown algae, such as fucoidan, alginate oligosaccharide, phlorotannins, and fucoxanthin, have demonstrated significant pharmaceutical utility[14]. Of these, fucoidan exhibits a variety of bioactivities, such as anti-cancer, antitumor, anti-thrombotic, anti-inflammatory, antivirus, immunomodulatory, hepatoprotective, and neuroprotective effects[15–19]. Fucoidan has been clinically used for nearly two decades to treat renal diseases[20]. As a fucoidan-rich brown alga, *S. sculpera* is an ideal candidate for investigating the mechanisms of fucoidan accumulation in kelp, with research on this species promising to deliver critical insights into fucoidan biosynthesis pathways. Thus, deciphering the genome of *S. sculpera* is essential for revealing the genetic foundation underlying its fucoidan synthesis and accumulation mechanisms.

[1]Ocean School, Yantai University, Yantai, 264005, China. [2]Shandong Oriental Ocean Sci-Tech Co., Ltd., Yantai, 264003, China. [3]Provincial Key Laboratory of Marine Seed Industry of Shandong, Yantai, 264003, China. [4]Shandong Technology Innovation Center of Algae and Sea Cucumber, Yantai, 264003, China. [5]These authors jointly supervised this work: Xiaojie Li, Quansheng Zhang. ✉e-mail: yeslxj@sina.com; ytuqsz@hotmail.com

Library Seqeuncing	Clean Data (Gb)	Coverage (X)
Illumina	80.52	95
PacBio	26.83	31
Hi-C	75.75	90
RNA	8.19	9

Table 1. Sequencing data of *S. sculpera*.

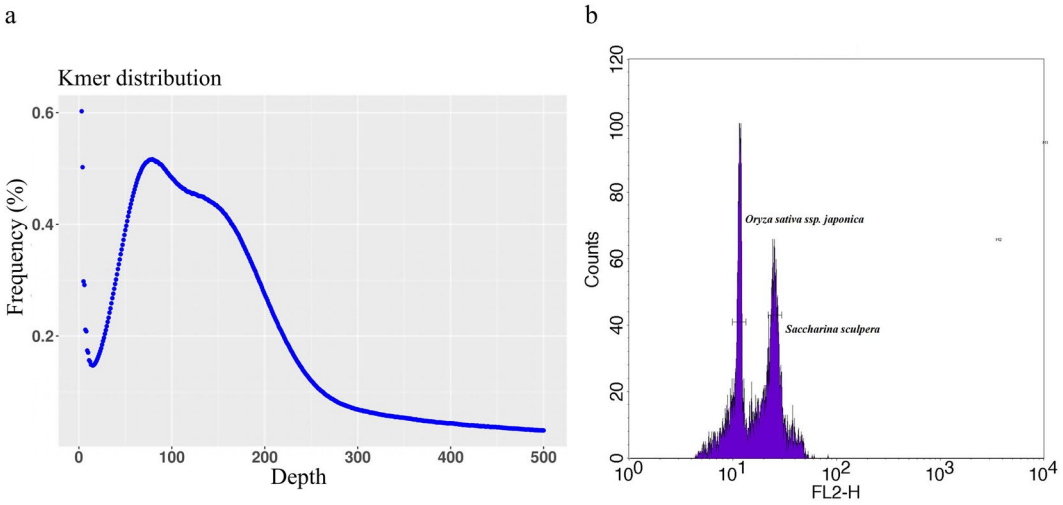

Fig. 1 The genome survey of *S. sculpera*. (**a**) The 19-Kmer analysis to estimate the genome size. (**b**) Estimation using flow cytometry techniques.

Phaeophyceae, a class of brown algae, encompass approximately 2000 species, which are taxonomically classified into 19 orders, 59 families, and 340 genera[21]. Since the publication of the first nuclear genome of brown algae (*Ectocarpus siliculosus*) in 2010, a total of 59 brown algal genomes have been archived in the National Center for Biotechnology Information, of which six have chromosome-level assemblies. Of the 59 brown algal genomes, 17 belong to the Laminariaceae family, of which three have chromosome-level assemblies. Compared to the genomes of terrestrial higher plants, the genomes of brown algae are more challenging to map due to difficulties in nucleic acid extraction, genome complexity, and their unique evolutionary history[22]. Nucleic acid extraction of brown algae is complicated by its high polysaccharide and polyphenol contents. Additionally, the genome of brown algae is highly complex, with extensive repetitive sequences, which limit its high-quality chromosome-level assembly. Lastly, gene annotation is challenging due to the phylogenetic distance between brown algae and other well-characterised eukaryotic lineages (e.g., animals, plants, and fungi).

In this research, we assembled a chromosome-level genome of *S. sculpera* through the integration of Illumina short-reads, PacBio HiFi, and high-throughput chromosome conformation capture (Hi-C) data. This strategy yielded a 726.56 Mb genome assembly, 98% of which was anchored to 34 chromosomes, with a scaffold N50 of 19.12 Mb. Gene prediction revealed that repetitive sequences constituted 60.26% of the genome and that 22,197 of the 25,685 (86.42%) protein-coding genes function were annotated. The assessment metrics of the genome assembly, including the Merqury quality value (37.69), the completeness (92.51%), and paired-end short-read alignment rate (97.30%), suggest a high degree of genome completeness. This research provides essential foundational data and valuable genetic resources for understanding the mechanisms underlying fucoidan biosynthesis in *S. sculpera*, as well as for advancing the understanding of brown algal ecology and evolution.

Methods

Sample collection and nucleic acid extraction. *S. sculpera* were originally obtained from Japan in December 2012 and subsequently cultivated in Rongcheng, Shandong Province, China, by Shandong Oriental Ocean Sci-Tech Co., Ltd., an enterprise specializing in marine species hatchery and mariculture. Juvenile *S. sculpera* sporophytes (~10 cm) were collected from the artificial raft culture experimental area in Rongcheng (122.08°E, 36.45°N), Shandong Province, China. The samples were repeatedly scrubbed with sterile absorbent cotton soaked in sterilized seawater, immersed in a 1.5% potassium iodide solution for 15 min, and treated with an 80–100 mg/L polymyxin–penicillin solution for 15 min. The purified samples were immediately frozen in liquid nitrogen and stored at −80 °C. High-quality genomic DNA was extracted from the *S. sculpera* blades via the cetyltrimethylammonium bromide method[23], and the total RNA was extracted using the Polysaccharide Polyphenol Plant Total RNA Extraction Kit (DP441, TianGen, China).

Library construction and sequencing. Genomic DNA was randomly fragmented by ultrasonic sonication to produce target-sized fragments (350 bp), which were subjected to end repair, adenine tailing, and adapter

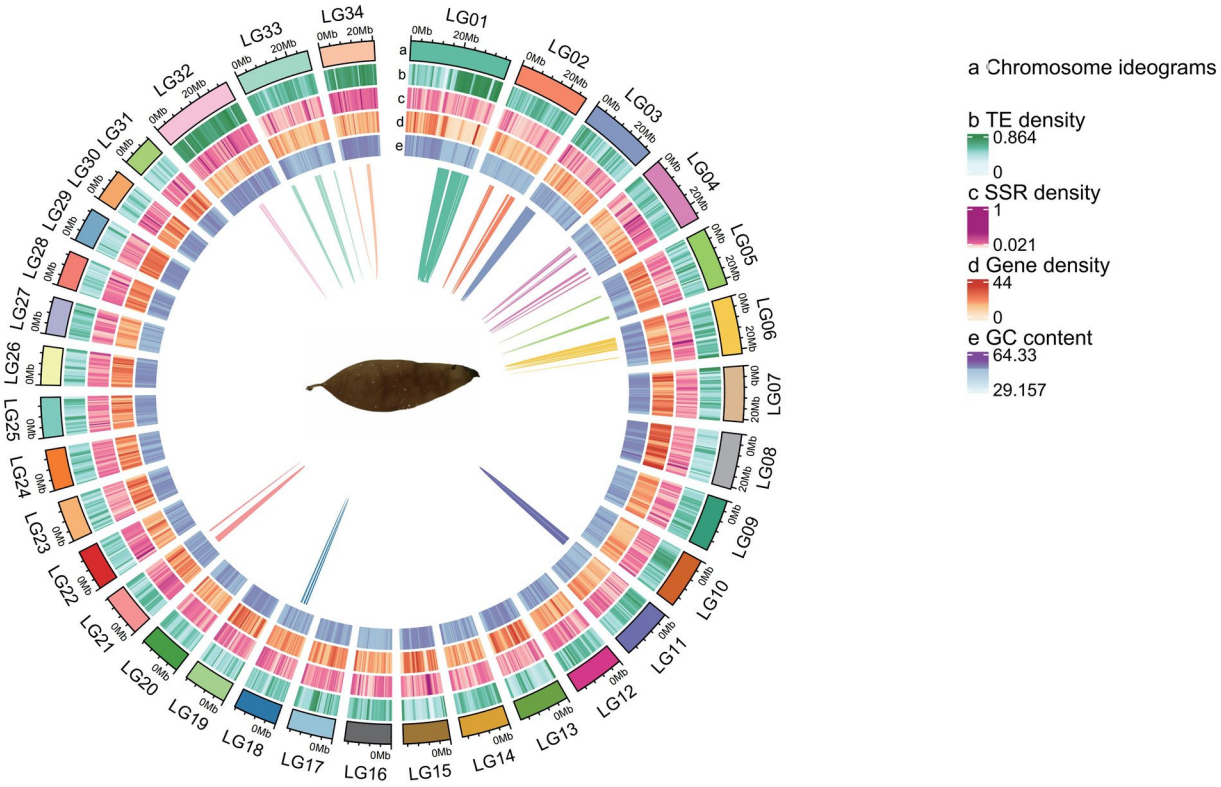

Fig. 2 The feature of *S. sculpera* genome and the juvenile sporophyte of *S. sculpera*.

ligation, size selection (350 ± 50 bp), and PCR amplification to construct short-read libraries. RNA libraries were prepared by enriching mRNA using Oligo(dT) beads, fragmenting the mRNA, synthesizing double-stranded cDNA, repairing ends, ligating adapters, and amplifying by PCR. Next, sequencing of the libraries was carried out on the Illumina NovaSeq. 6000 platform, which produced 80.52 Gb of clean data, covering 95.32 × of the *S. sculpera* genome and approximately 8.19 Gb of RNA-sequencing data. The single-molecule real-time PacBio libraries were prepared utilizing the SMRTbell® express template prep kit 2.0 (PacBio, CA, USA); they were sequenced using the circular consensus sequencing (CCS) mode on the PacBio Sequel II platform to generate totally 26.83 Gb clean HiFi data, covering approximately 31 × of the *S. sculpera* genome. Hi-C library was constructed through the *in situ* Hi-C method, involving cell crosslinking, restriction enzyme digestion, end repair, ligation, DNA purification, capture, and sequencing, generating 75.75 Gb sequencing data (Table 1).

Genome survey. We performed a k-mer frequency analysis (k = 19) utilizing 80.52 Gb Illumina clean data by Jellyfish (v 2.1.4)[24] to characterize the features of *S. sculpera* genome. According to the distribution of 19-kmers, the *S. sculpera* genome was estimated to be 844.73 Mb in size, with 0.44% heterozygosity and 76.71% repeat content (Fig. 1a). Moreover, flow cytometry analysis (BD FACScalibur flow cytometer, BD Biosciences) using *Oryza sativa* ssp. *japonica* genome (430 Mb) as a reference revealed that the genome size of *S. sculpera* was approximately 890 Mb, consistent with the K-mer assessment (Fig. 1b).

Genome assembly. The 26.83 Gb PacBio HiFi reads were used to generate a contig-level genome using Hifiasm (v0.16)[25]. For the anchored contigs, 75.75 Gb clean Hi-C data were mapped to the polished genome by BWA (v0.7.17-r1188)[26]. The adaptors and low-quality Hi-C reads were filtered via HiC-Pro (v2.10.0)[27]. The contigs were clustered, ordered, and oriented onto chromosomes using LACHESIS (v2.0.1)[28] with default parameters. The chromosome-scale resolution of the genome assembly was achieved through Hi-C scaffolding, complemented by manual curation of the interaction heatmap. A total of 711.82 Mb of assembled sequences were anchored onto 34 chromosomes (Fig. 2). Gap filling was performed on chromosomal sequences using the CCS data, with the sequential application of gap-closing tools, TGS_gapcloser and LR_Gapcloser, for refinement. The chromosome-scale genome of *S. sculpera* was 726.56 Mb, with a Scaffold N50 of 19.12 Mb, representing about 98% of the draft genome, which indicates high completeness of the chromosomal organization (Fig. 3 & Table 2)

Repeat annotation. Transposon elements (TEs) were predicted by homology searching and *de novo* methods. Firstly, the *de novo* repeat library of the *S. sculpera* genome was generated by RepeatModeler v2.0.1 (parameters: BuildDatabase -name & RepeatModeler -pa 12)[29] using RepeatScout[30] and RECON[31]. Thereafter, full-length long terminal repeat retrotransposons (fl-LTR-RTs) were identified by LTRharvest (v1.5.9)[32] and LTR_finder v1.07 (parameters: ltr_finder -w 2 -C -D)[33]. Subsequently, both fl-LTR-RT and non-redundant LTR libraries could be constructed using LTR_retriever (v2.9.0)[34]. Then *de novo* TE sequence library was combined with the

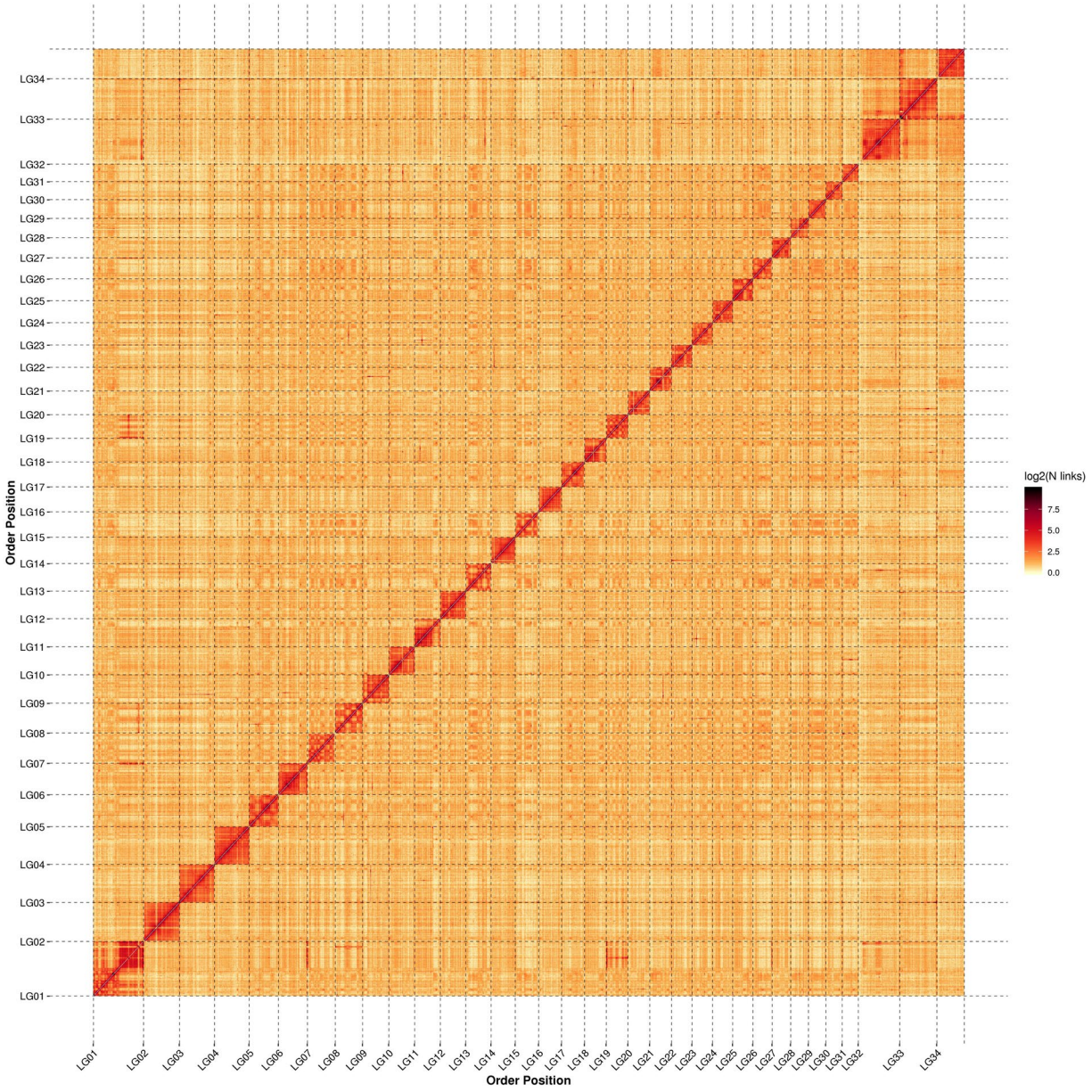

Fig. 3 The interactive heatmap of Hi-C assembled chromosomes, in which LG01-34 are the chromosome numbers 1–34. The colour spanning from yellow to deep represents the interaction intensity from low to high.

well-established Dfam (v3.5) database to construct a non-redundant species-specific TE library. A homology search was then performed against this library using RepeatMasker v4.1.0 (parameters: repeatmasker -nolow -no_is -norna -engine wublast -parallel 8 -qq)[35] to identify and classify the final TE sequences in the *S. sculpera* genome. Tandem repeats were annotated utilizing Tandem Repeats Finder v409 (parameters: trf 2 7 7 80 10 50 500 -d -h)[36] and the MIcroSAtellite (v2.1)[37]. Repetitive sequences were found to constitute 60.26% (438.5 Mb) of the total genome, of which 51.06% (371.56 Mb) was TEs, including RTs (63%) like LTR/Gypsy, LINE, and LTR/Copia families and 9.2% (66.94 Mb) were tandem repeats, including the microsatellites, minisatellites, and satellites (Table 3).

Gene prediction and functional annotation. *De novo*, homology-based, and transcriptome-based predictions were utilised to predict protein-coding genes in the *S. sculpera* genome. Augustus (v3.1.0)[38] and SNAP (2006-07-28)[39] were employed to predict the *de novo* gene models. GeMoMa v1.7 (parameters: run.sh mmseqs)[40] was used to perform homology method alignment with reference gene models from *Ectocarpus siliculosus*, *Nemacystus decipiens*, *Saccharina fusiforme*, *Saccharina japonica*, *Undaria pinnatifida Kr2015*, and *Undaria pinnatifida M23*. The annotation pipeline employed HISAT2 v2.1.0 (parameters: hisat2 --dta -p 10)[41] for RNA-seq read alignment, StringTie2 v2.1.4 (parameters: stringtie -p 2)[42] for transcript assembly, and GeneMarkS-T (v5.1)[43] to predict the evidence-driven gene model, ensuring the integration of transcriptomic data into structural annotation. Gene structural annotation was conducted using PASA (v2.4.1)[44] with transcript evidence derived from

Chromosome ID	Cluster Num	Cluster Len	Order Num	Order Len
LG01	222	39,962,978	143	37,782,919
LG02	104	28,601,561	78	26,892,317
LG03	135	31,251,629	92	26,266,750
LG04	142	28,628,393	96	25,827,640
LG05	215	25,658,280	125	22,145,021
LG06	180	22,617,856	121	21,691,127
LG07	226	22,342,778	133	20,831,162
LG08	286	22,694,946	175	20,671,743
LG09	135	20,366,086	90	19,681,142
LG10	158	20,319,968	102	19,481,105
LG11	158	21,679,176	106	19,449,568
LG12	137	21,920,209	97	19,145,279
LG13	238	21,249,014	141	19,034,598
LG14	123	18,646,978	88	18,021,967
LG15	219	19,962,887	148	17,473,232
LG16	121	19,490,696	93	17,332,965
LG17	213	18,358,066	139	17,224,478
LG18	144	18,036,630	103	16,437,605
LG19	241	17,941,500	144	16,390,716
LG20	161	18,054,706	118	16,256,577
LG21	155	17,000,655	102	16,213,116
LG22	155	16,406,665	104	15,467,416
LG23	188	17,200,886	114	15,433,617
LG24	182	16,458,877	105	15,276,027
LG25	169	16,194,772	104	15,153,190
LG26	181	15,407,317	127	14,550,101
LG27	127	15,301,304	87	13,960,671
LG28	171	15,466,302	104	13,374,879
LG29	209	14,148,954	134	12,997,640
LG30	174	13,451,062	107	12,355,756
LG31	178	13,110,287	112	12,012,792
LG32	92	31,589,193	68	31,096,716
LG33	114	31,757,297	80	27,741,303
LG34	37	20,546,214	36	20,528,133
Total(Ratio %)	90.56	98.31	65.31	91.90

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

Trinity v2.11 (parameters: Trinity--genome_guided_bam)[45] assembled unigenes. Gene models from three distinct methodologies were integrated via the EVM (v1.1.1)[46] and then improved and polished using PASA. The prediction of gene functions was carried out according to the alignments on multiple databases, including the NR, EggNOG[47], KOG, TrEMBL, Pfam[48], KEGG[49], and Swiss-Prot[50] protein databases, utilizing diamond blastp (v0.9.29.130) with an E-value threshold of 1E-3. A total of 25,685 protein-coding genes were successfully identified, of which 22,197 (86.42%) were functionally annotated (Table 4). Lastly, we identified 1,719 tRNAs by tRNAscan-SE (v1.3.1)[51]; 1,124 rRNAs using Barrnap (v0.9)[52]; and 8 miRNAs, 24 snoRNAs, and 65 snRNAs using Infernal (v1.1)[53] against the Rfam database (v14.5)[54].

Data Records

The complete dataset for *S. sculpera*, including raw sequencing data and the assembled genome, is available through the National Center for Biotechnology Information (NCBI) and figshare. The raw data comprising Illumina, two PacBio, Hi-C, and RNA sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under accession numbers SRR33206832[55], SRR33206831[56], SRR33206830[57], SRR33206829[58], and SRR33206828[59]. The genome assembly has been deposited in NCBI GenBank under the WGS accession number JBQGVP000000000[60]. Additionally, the chromosome-level genome assembly of *S. sculpera*, along with its predicted coding sequences and protein sequences, is available at figshare[61].

Technical Validation

Low contamination ratio. The high-quality assembly of *S. sculpera* genome was dependent on its minimal contamination during sample treatment. First, to assess potential DNA contamination, we randomly selected 10,000 single-ended reads and blasted them against the NCBI NT database. BLAST analysis against the NT database assigned the reads to multiple species, including three species of the *Saccharina* genus, and the remaining

Type	Number	Length	Rate (%)
ClassI:Retroelement	677,391	234,223,482	32.18
ClassI/DIRS	5,613	4,968,612	0.68
ClassI/LINE	124,666	32,124,330	4.41
ClassI/LTR/Copia	35,523	18,663,942	2.56
ClassI/LTR/ERV	2,100	224,953	0.03
ClassI/LTR/Gypsy	60,014	42,873,969	5.89
ClassI/LTR/Ngaro	9,235	6,322,959	0.87
ClassI/LTR/Pao	259	37,908	0.01
ClassI/LTR/Unknown	430,155	127,742,622	17.55
ClassI/SINE	9,826	1,264,187	0.17
ClassII:DNA transposon	475,410	137,344,825	18.87
ClassII/CACTA	365	28,191	0
ClassII/Crypton	2	111	0
ClassII/Dada	382	189,043	0.03
ClassII/Ginger	875	100,805	0.01
ClassII/Helitron	574	233,892	0.03
ClassII/IS3EU	72	6,830	0
ClassII/Kolobok	170	24,640	0
ClassII/Maverick	136	16,176	0
ClassII/Merlin	31	1,946	0
ClassII/Mutator	136	12,541	0
ClassII/P	20	1,291	0
ClassII/PIF-Harbinger	1,999	381,126	0.05
ClassII/PiggyBac	11	784	0
ClassII/Sola	1	98	0
ClassII/Tc1-Mariner	70	4,685	0
ClassII/Unknown	467,355	135,686,721	18.64
ClassII/Zator	236	82,580	0.01
ClassII/Zisupton	20	1,168	0
ClassII/hAT	2,955	572,197	0.08
Unknown	8	718	0
microsatellite	432,228	12,910,114	1.77
minisatellite	229,088	23,443,915	3.22
satellite	35,098	30,627,768	4.21
Total	1,849,223	438,550,822	60.26

Table 3. Statistics of repeat elements in *S. sculpera* genome. The calculation results of the rate should be rounded to two decimal places.

Anno_Database	Annotated_Number	Annotated_Ratio (%)
GO	12,258	47.72
KEGG	8,453	32.91
KOG	8,014	31.20
Pfam	14,392	56.03
Swissprot	10,322	40.19
TrEMBL	21,804	84.89
eggNOG	9,101	35.43
NR	19,121	74.44
All	22,197	86.42

Table 4. Functional annotation in *S. sculpera* genome.

species each accounted for less than 1% of aligned reads, suggesting negligible bacterial contamination in this project (Table 5). Moreover, to evaluate potential contamination from non-target species, we aligned the HiFi data to the genome utilizing BWA (v0.7.17). The genome was then partitioned into consecutive, non-overlapping 10-kb windows (with terminal regions shorter than 10 kb retained as their actual lengths). For each window, the mean sequencing depth and GC content percentage were calculated to assess the correlation between GC content distribution and sequencing depth. The results showed that GC content was distributed within the 45–55% range,

Species	Aligned percentage (%)
Saccharina latissima	2.44
Saccharina japonica	2.21
Saccharina sculpera	0.8
Other	2.18
Total	7.63

Table 5. Species of reads annotated to NT database.

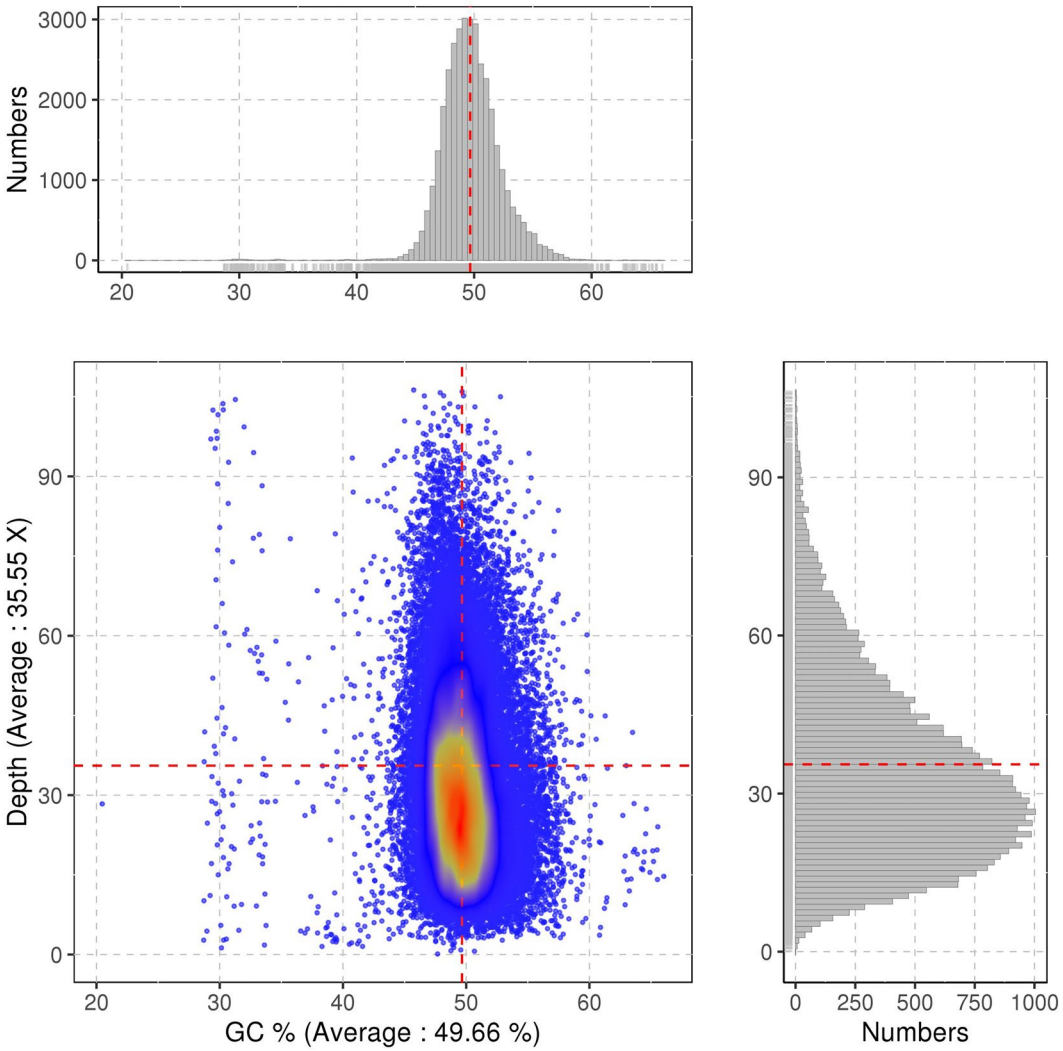

Fig. 4 Distribution of GC content and sequencing depth. The abscissa denotes GC content, and the ordinate shows sequencing depth. The right and top sides show the distribution of sequencing depth and GC content, respectively.

and the sequencing depth was concentrated in the 25–40X region, indicating that the genomic samples were free from contamination (Fig. 4).

Genome assembly evaluation. The quality of the genome was evaluated using multiple metrics. To evaluate the completeness of the assembly and consistency of sequencing coverage, we aligned illumina sequencing data and PacBio CCS data to the genome by BWA and Minimap2[62]. It has shown that the proportion of mapped short reads and HiFi reads were 97.30% and 97.78%, with genome coverage of 99.85% and 99.69%, respectively. The integrity and quality of the assembled genome were evaluated using Merqury (v1.3)[63], yielding a consensus quality value of 37.69 and k-mer completeness of 92.51%. Then, the Hi-C heatmap (Fig. 3) demonstrated strong interaction signals along the diagonals across all chromosomes, verifying proper contig ordering and orientation, thereby validating the continuity and accuracy of the genome assembly. In addition, we performed genome-wide

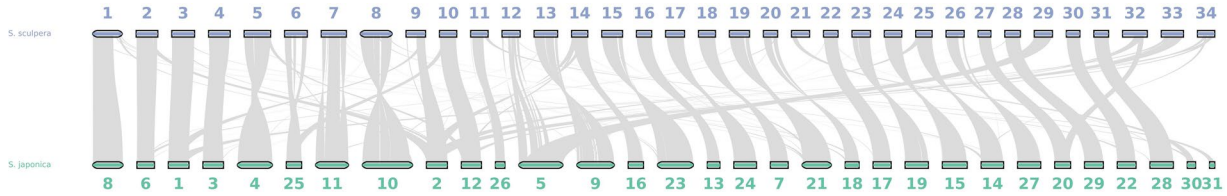

Fig. 5 Collinearity analysis between *S. sculpera* and *S. japonica*.

collinearity analysis between *S. sculpera* and its congener *S. japonica*. Homologous gene pairs were identified using DIAMOND with stringent criteria (e-value less than $1e^{-5}$, coverage score greater than 0.5). Collinear blocks were determined via MCScanX[64], and results visualized with JCVI (v0.9.13)[65]. Comparative analysis revealed strong collinearity between *S. sculpera* and *S. japonica*, confirming the high-quality sequencing and assembly of the *S. sculpera* genome (Fig. 5). Altogether, these outcomes demonstrate the *S. sculpera* genome assembly exhibits high levels of completeness, quality, and correctness.

Data availability

The data associated with *S. sculpera* are available through the NCBI and figshare. The raw sequencing data (Illumina, PacBio, Hi-C, and RNA-seq) generated in this study are available in the NCBI SRA under the accession numbers SRR33206832, SRR33206831, SRR33206830, SRR33206829, and SRR33206828. The whole genome shotgun project has been deposited at GenBank under the accession number JBQGVP000000000. The chromosome-level genome assembly, annotated coding sequences, and protein sequences are available on figshare (10.6084/m9.figshare.28831976).

Code availability

In this study, the guidelines of all bioinformatics tools used were strictly followed. No custom code was employed, and all operations were carried out based on the software protocols. These software used is publicly available. For specific versions and parameter details, please refer to the Methods of this paper.

Received: 22 April 2025; Accepted: 24 September 2025;
Published online: 07 November 2025

References

1. Kim, S. H., Kim, Y. D., Hwang, M. S., Hwang, E. K. & Yoo, H. I. Temperature ranges for survival and growth of juvenile *Saccharina sculpera* (Laminariales, Phaeophyta) and applications for field cultivation. *Algae* **36**, 231–240, https://doi.org/10.4490/algae.2021.36.11.21 (2021).
2. Sato, Y., Kozono, J., Nishihara, G. N. & Terada, R. Effect of light and temperature on photosynthesis of a cultivated brown alga, *Saccharina sculpera* (Laminariales), from Japan. *Phycologia* **59**, 375–384, https://doi.org/10.1080/00318884.2020.1777384 (2020).
3. Tani, T., Kawagoe, C., Matsumoto, S., Mizuta, H. & Yasui, H. Seasonal variations and morphological changes of sporophytes of *Saccharina sculpera* (Laminariales, Phaeophyceae) from Hakodate, Hokkaido. *Aquac. Sci.* **63**, 235–244, https://doi.org/10.11233/aquaculturesci.63.235 (2015).
4. Tang, S. *et al*. Enzyme-assisted extraction of fucoidan from *Kjellmaniella crassifolia* based on kinetic study of enzymatic hydrolysis of algal cellulose. *Algal Res.* **66**, 102795, https://doi.org/10.1016/j.algal.2022.102795 (2022).
5. Li, X. *et al*. Chromosome-level genome assembly for the ecologically and economically important alga *Saccharina japonica*. *Sci. Data* **12**, 290, https://doi.org/10.1038/s41597-025-04620-w (2025).
6. Liu, S. *et al*. Studies on the hepatoprotective effect of fucoidans from brown algae *Kjellmaniella crassifolia*. *Carbohydr. Polym.* **193**, 298–306, https://doi.org/10.1016/j.carbpol.2018.03.077 (2018).
7. Song, Y. *et al*. The positive effects of fucoidans extracted from the brown seaweed *Saccharina japonica* on protection against CCl_4-induced liver injury. *J. Appl. Phycol.* **29**, 2077–2087, https://doi.org/10.1007/s10811-017-1097-3 (2017).
8. Dhakal, S., Jüterbock, A. O., Lei, X. & Khanal, P. Application of the brown macroalga *Saccharina latissima* (Laminariales, Phaeophyceae) as a feed ingredient for livestock: A review. *Anim. Nutr.* **19**, 153–165, https://doi.org/10.1016/j.aninu.2024.07.001 (2024).
9. FAO. *The state of world fisheries and aquaculture 2024*. (Food and Agriculture Organization, 2024).
10. Rogel-Castillo, C., Latorre-Castañeda, M., Muñoz-Muñoz, C. & Agurto-Muñoz, C. Seaweeds in food: Current trends. *Plants* **12**, 2287, https://doi.org/10.3390/plants12122287 (2023).
11. Yang, Z., Li, X., Yu, M., Jiang, S. & Qi, H. Effects of different processing methods on the quality and physicochemical characteristics of *Laminaria japonica*. *Foods* **12**, 1619, https://doi.org/10.3390/foods12081619 (2023).
12. Janarthanan, M. & Senthil Kumar, M. The properties of bioactive substances obtained from seaweeds and their applications in textile industries. *J. Ind. Text.* **48**, 361–401, https://doi.org/10.1177/1528083717692596 (2018).
13. Polat, S. *et al*. Recent advances in industrial applications of seaweeds. *Crit. Rev. Food Sci. Nutr.* **63**, 4979–5008, https://doi.org/10.1080/10408398.2021.2010646 (2023).
14. Anjana, K. & Arunkumar, K. Brown algae biomass for fucoxanthin, fucoidan and alginate; Update review on structure, biosynthesis, biological activities and extraction valorisation. *Int. J. Biol. Macromol.* **280**, 135632, https://doi.org/10.1016/j.ijbiomac.2024.135632 (2024).
15. Nishino, T., Nagumo, T., Kiyohara, H. & Yamada, H. Structural characterization of a new anticoagulant fucan sulfate from the brown seaweed *Ecklonia kurome*. *Carbohydr. Res.* **211**, 77–90, https://doi.org/10.1016/0008-6215(91)84147-7 (1991).
16. Han, Y.-S., Lee, J. H., Chang, H. S. & Lee, S. H. Inhibition of Lewis lung cancer cell growth and migration by fucoidan. *Mol. Cell. Toxicol.* **10**, 269–276, https://doi.org/10.1007/s13273-014-0030-y (2014).
17. Zaporozhets, T. S., Ermakova, S. V., Zvyagintseva, T. N. & Besednova, N. N. Antitumor effects of sulfated polysaccharides produced from marine algae. *Biol. Bull. Rev.* **4**, 122–132, https://doi.org/10.1134/S2079086414020078 (2014).
18. Yang, J. W., Yoon, S. Y., Oh, S. J., Kim, S. K. & Kang, K. W. Bifunctional effects of fucoidan on the expression of inducible nitric oxide synthase. *Biochem. Biophys. Res. Commun.* **346**, 345–350, https://doi.org/10.1016/j.bbrc.2006.05.135 (2006).

19. Berteau, O. Sulfated fucans, fresh perspectives: Structures, functions, and biological properties of sulfated fucans and an overview of enzymes active toward this class of polysaccharide. *Glycobiology* **13**, 29R–40, https://doi.org/10.1093/glycob/cwg058 (2003).
20. Lin, Z. *et al*. Molecular targets and related biologic activities of fucoidan: A review. *Mar. Drugs* **18**, 376, https://doi.org/10.3390/md18080376 (2020).
21. Akita, S. *et al*. Providing a phylogenetic framework for trait-based analyses in brown algae: Phylogenomic tree inferred from 32 nuclear protein-coding sequences. *Mol. Phylogenet. Evol.* **168**, 107408, https://doi.org/10.1016/j.ympev.2022.107408 (2022).
22. Cock, J. M. *et al*. Genomics of brown algae: Current advances and future prospects. *Genes Genomics* **34**, 1–5, https://doi.org/10.1007/s13258-010-0225-0 (2012).
23. Sahu, S. K., Thangaraj, M. & Kathiresan, K. DNA extraction protocol for plants with high levels of secondary metabolites and polysaccharides without using liquid nitrogen and phenol. *ISRN Mol. Biol.* **2012**, 205049, https://doi.org/10.5402/2012/205049 (2012).
24. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770, https://doi.org/10.1093/bioinformatics/btr011 (2011).
25. 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, https://doi.org/10.1038/s41592-020-01056-5 (2021).
26. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760, https://doi.org/10.1093/bioinformatics/btp324 (2009).
27. Servant, N. *et al*. HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259, https://doi.org/10.1186/s13059-015-0831-x (2015).
28. Burton, J. N. *et al*. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125, https://doi.org/10.1038/nbt.2727 (2013).
29. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* **117**, 9451–9457, https://doi.org/10.1073/pnas.1921046117 (2020).
30. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358, https://doi.org/10.1093/bioinformatics/bti1018 (2005).
31. Bao, Z. & Eddy, S. R. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* **12**, 1269–1276, https://doi.org/10.1101/gr.88502 (2002).
32. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18, https://doi.org/10.1186/1471-2105-9-18 (2008).
33. Xu, Z. & Wang, H. LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268, https://doi.org/10.1093/nar/gkm286 (2007).
34. Ou, S. & Jiang, N. LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422, https://doi.org/10.1104/pp.17.01310 (2018).
35. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma.* **25**, 4.10.1–4.10.14, https://doi.org/10.1002/0471250953.bi0410s25 (2009).
36. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580, https://doi.org/10.1093/nar/27.2.573 (1999).
37. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **33**, 2583–2585, https://doi.org/10.1093/bioinformatics/btx198 (2017).
38. 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, https://doi.org/10.1093/bioinformatics/btn013 (2008).
39. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59, https://doi.org/10.1186/1471-2105-5-59 (2004).
40. Keilwagen, J. *et al*. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89–e89, https://doi.org/10.1093/nar/gkw092 (2016).
41. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360, https://doi.org/10.1038/nmeth.3317 (2015).
42. Pertea, M. *et al*. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295, https://doi.org/10.1038/nbt.3122 (2015).
43. Tang, S., Lomsadze, A. & Borodovsky, M. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res.* **43**, e78–e78, https://doi.org/10.1093/nar/gkv227 (2015).
44. Haas, B. J. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666, https://doi.org/10.1093/nar/gkg770 (2003).
45. Grabherr, M. G. *et al*. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652, https://doi.org/10.1038/nbt.1883 (2011).
46. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7, https://doi.org/10.1186/gb-2008-9-1-r7 (2008).
47. 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, https://doi.org/10.1093/nar/gky1085 (2019).
48. Finn, R. D. Pfam: Clans, web tools and services. *Nucleic Acids Res.* **34**, D247–D251, https://doi.org/10.1093/nar/gkj149 (2006).
49. 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, https://doi.org/10.1093/nar/gkv1070 (2016).
50. Boeckmann, B. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370, https://doi.org/10.1093/nar/gkg095 (2003).
51. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964, https://doi.org/10.1093/nar/25.5.955 (1997).
52. Loman, T. A novel method for predicting ribosomal RNA genes in prokaryotic genomes. (Lund University, 2017).
53. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935, https://doi.org/10.1093/bioinformatics/btt509 (2013).
54. Griffiths-Jones, S. Rfam: Annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124, https://doi.org/10.1093/nar/gki081 (2004).
55. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR33206832 (2025).
56. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR33206831 (2025).
57. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR33206830 (2025).
58. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR33206829 (2025).
59. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR33206828 (2025).
60. *NCBI GenBank* https://identifiers.org/ncbi/insdc:JBQGVP000000000 (2025).
61. Guo, L. Chromosome-scale genome assembly and annotation of *Saccharina sculpera*. *Figshare* https://doi.org/10.6084/m9.figshare.28831976 (2025).
62. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100, https://doi.org/10.1093/bioinformatics/bty191 (2018).

63. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245, https://doi.org/10.1186/s13059-020-02134-9 (2020).
64. Wang, Y. *et al*. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**(7), e49–e49, https://doi.org/10.1093/nar/gkr1293 (2012).
65. Tang, H. *et al*. jcvi: JCVI utility libraries. *Zenodo* https://doi.org/10.5281/zenodo.31631 (2015).

Acknowledgements

This reaserch was supported by the National Natural Science Foundation of China (No. 42476112) and the earmarked fund for CARS (CARS-50).

Author contributions

Q.S.Z. and X.J.L. conceived and designed the research, acquired funding, and participated in the writing, review and editing. M.Y.Z. and W.W.W. performed the experiments and analyzed the data. L.H.G. analyzed the data and composed the initial manuscript. All authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to X.L. or Q.Z.