




Structure, sequence and phylogenetic analysis of the complete chloroplast genome of the brown algae *Saccharina sp. ye-E* (Laminariaceae: Phaeophyceae) from Sakhalin Oblast, Russia

Wei Zhang^{1,2} , Ziwen Liu^{1,2} , Xiao Fan² ^a

¹ Graduate School of Chinese Academy of Agricultural Sciences, Beijing, China, ² Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, Shangdong, China

Keywords: complete chloroplast genome, Illumina sequencing, *Saccharina sp. ye-E*

<https://doi.org/10.46989/001c.77614>

Israeli Journal of Aquaculture - Bamidgeh

Vol. 75, Issue 1, 2023

For this investigation, the chloroplast genome of *Saccharina sp. ye-E* from Russia was uncovered and annotated with Illumina sequencing data to examine the phylogenetic relationship of *Saccharina* in Laminariaceae from a molecular biology perspective. Analysis of the structural characteristics, simple repeat sequence (SSR) loci, relative species differences, codon preference, and phylogenetic relationships was conducted. The results revealed that the chloroplast genome of *Saccharina sp. ye-E* had a length of 130,624 bp, containing 139 protein-coding genes (PCGs), 6 ribosomal RNAs (rRNAs) and 57 transfer RNAs (tRNAs) genes, and a total GC content of 31.1%. There were 15 SSR loci in the genome. Effective codon number (ENC) and codon adaptation index (CAI) both indicated the strong codon randomness and codon preference. The phylogenetic tree, based on the complete chloroplast genomes of 10 brown algae, showed that four species of *Saccharina* genus formed a clade, with *Saccharina sp. ye-E* and *Saccharina latissimi* having the most related genetic affinity. It is believed that the determination of the chloroplast genome of *Saccharina sp. ye-E* will be beneficial for future algal genetics evolution and systematic studies in the Laminariaceae.

INTRODUCTION

Saccharina, commonly known as kelp, is a genus of large brown algae in the family Laminariaceae. *Saccharina* species are native to many parts of the world, including the North Pacific, North Atlantic, and Southern Ocean. In addition to providing a habitat for various marine organisms, *Saccharina* species are an essential source of food for many marine organisms and have a range of ecological and economic uses. From an ecological standpoint, *Saccharina* plays a vital role in forming kelp forests, which absorb carbon dioxide through photosynthesis and provide a habitat for fish and other animals. From an economic standpoint, *Saccharina* is a major cultivated seaweed species, used not only as food but also as the primary raw material for extracting alginates, fucoidan, and mannitol. It is also highly valued in China, Japan, and Korea due to its commercial use as a raw material in food and industry.^{1,2}

The degradation of *Saccharina* kelp forests is one of the most pressing threats to coastal ecosystems. Loss and decline of kelp forests have been reported in temperate, boreal, and arctic regions. They are mainly attributed to an-

thropogenic drivers such as climate change, eutrophication, overfishing, and habitat destruction.³ To understand and protect these critical ecosystems, phylogenetic research is urgently needed to identify genetic diversity among species, define evolutionary lineages, and elucidate the group's evolutionary history. Such research can help us understand the evolutionary origins of this species, its current state of decline, and its potential for future recovery. With increased knowledge of the phylogeny of kelp species, it may be possible to develop conservation strategies to protect them in their current locations and restore populations in areas where they have been over-harvested or otherwise impacted. Additionally, phylogenetic studies can inform our understanding of how kelp species will respond to climate change and other environmental pressures in the future.⁴ Chloroplast DNA (cpDNA) is the most widely used marker for phylogenetic studies due to its high mutation rate, fast evolutionary rate, and maternal inheritance. cpDNA is a unique form of genetic material that is found in the chloroplastid of all photosynthetic organisms. As a result, cpDNA has been used in genetic studies to trace maternal ancestry and study species' population structure. In plants, cpDNA studies have been used to un-

a * Corresponding author. e-mail: fanxiao@ysfri.ac.cn.

derstand phylogeny and evolution and assess genetic diversity in specified.⁵

The application of cpDNA in the genus *Saccharina* has been particularly useful for understanding the population structure of the species.⁶ Using cpDNA markers, researchers have identified different genetic lineages within the genus, which can help inform conservation and management decisions. For example, cpDNA analysis has revealed that different populations of *Saccharina japonica* have distinct evolutionary histories and that the species is composed of two distinct genetic lineages.⁶ This finding has significant implications for the conservation and management of the species, as different populations may require different management strategies. We analyzed the structural characteristics, simple repeat sequence (SSR) loci, relative species differences, codon preference and phylogenetic relationships in this study to determine the complete chloroplast genome sequence of a wild strain of *Saccharina* collected from Russia to increase genetic variance and economic value of *japonica*.

MATERIALS AND METHODS

The *Saccharina sp. ye-E* sample used in the experiment was collected from the Sakhalin Oblast, Russia (50° 49'N, 142° 17'E). It was identified as *Saccharina latissimi* and labeled ye-E by the Yellow Sea Fisheries Research Institute of the Chinese Academy of Fishery Sciences. After the sample collection, it was cleaned with sterile water, dried with absorbent paper, frozen in liquid nitrogen and kept in an ultra-low temperature refrigerator for later use.

Obtain 0.1 g of *Saccharina sp. ye-E* samples and place them in a mortar that has been cooled using liquid nitrogen. After crushing, the genome was extracted using the CTAB method. Assess the purity and concentration of the DNA through 1% agarose gel electrophoresis and a multi-functional enzyme marker, and then submit it to a biological company for sequencing.

The paired-end reads were generated by using the Illumina sequencing system (Tianjin, China). Using Illumina sequencing, 152 million clean PE 100 reads of this DNA were obtained. The sequences were *de novo* assembled to contigs using ABySS Version2.0⁷ firstly and then linked by referring to the chloroplast genome *S. japonica* (NC018523, 130,584 in length) using BWA (Burrows-Wheeler Aligner). Of the total reads, 32% were mapped to the reference chloroplast genome *S. japonica*. No gaps were present in the mapping result. The consensus sequence was produced with Geneious version 11.1.4 (<http://www.genious.com>). Additionally, the chloroplast genome was annotated by GeSeq,⁸ AGORA,⁹ CPGAVAS,¹⁰ OGDRAW¹¹ MITOS server,¹² the tRNAscan-SE server¹³ and PGA,¹⁴ which was also verified manually. Finally, the complete chloroplast genome of *Saccharina sp. ye-E* was submitted to GenBank (accession NO. MZ706293.1).

Employ the MISA (<http://pgrc.ipk-gatersleben.de/misa/misa.html>) software to search for Simple Sequence Repeats (SSRs), with the following parameters: a mononucleotide repetition threshold of 10, and a dinucleotide repetition

threshold of 6; for trinucleotides, four nucleotides, five nucleotides and six nucleotides, the repetition threshold is 5. Utilize EMBOSS Explorer software (<https://www.bioinformatics.nl/emboss-explorer/>) to analyze the Effective Number of Codon (ENC), Relative Synonymous Codon Usage (RSCU), Codon Adaptation Index (CAI) and other pertinent parameters, where the RSCU value should be above 1 as the cut-off point, indicating that the codon is employed at a relatively high frequency; the ENC ranges from 21 to 60, and a value of 40 is the cut-off point for judging randomness in codons, while a value of 45 is the cut-off value for differentiating bias in codons. CAI values range from 0 to 1, with size proportional to codon preference.

Eight related species' chloroplast genome sequences were downloaded from NCBI database, namely *Laminaria rodriguezii* (MT732096.1), *Laminaria solidungula* (NC 044690.1), *Lessonia flavicans* (MN561187.1), *Macrocystis integrifolia* (MW899036.1), *Saccharina japonica* (NC 018523.1), *Saccharina sp. ye-B* (MW038824.1), *Saccharina latissima* (NC 049039.1), and *Undaria pinnatifida* (KP298002.1), with *Scytosiphon lomentaria* (MK798154.1) and *Ishige okamurae* (MW762687.1) as the outgroup. The multiple sequence alignments of the selected chloroplast genomes were built using MAFFT software, and the phylogenetic tree was constructed by applying Neighbor-Joining, MrBayes and Maximum likelihood optimality criteria using the concatenated nucleotide sequences shared among *Saccharina sp. ye-E* and the other 10 brown algae complete chloroplast genomes using IQtree2.¹⁵ All nodes were well supported by Bayesian posterior probabilities and 1000 bootstrap replicates.

RESULTS

The total length of the chloroplast genome sequence of *Saccharina sp. ye-E* is 130,624 bp (**Figure 1**). It consists of 139 protein-coding genes (PCGs), 3 ribosomal (rRNAs) genes (5S rRNA, 16S rRNA, 23S rRNA, two copies for each) and 57 transfer RNA (tRNAs) genes. These 139 functional genes can be divided into three categories: 39 related to photosynthesis, 71 involved in genetic replication systems, and 29 other genes. The total GC content of *Saccharina sp. ye-E* is 31.1%. The nucleotide frequency of the H-strand is as follows: T, 34.5%; A, 34.4%; C, 15.4%; and G, 15.7%. The chloroplast of *Saccharina sp. ye-E* encodes 32,407 amino acids with the start codons included. The two 5S rRNAs are 110 bp and 100 bp, respectively; the 16S rRNAs are both 1,480 bp; and the 23S rRNA genes are 2,944 bp and 2,946 bp in length, respectively. Chloroplast data supporting this study are openly available in GenBank at the nucleotide database, Associated Bio Project (PRJNA272647), BioSample (SAMN03740579), and SRA (SRS947870).

The chloroplast genome of *Saccharina sp. ye-E* contained 15 SSR markers, including three single nucleotides (A/T), which comprised 80% of the total. Additionally, two dinucleotides (AT/TA) and three trinucleotides (ATA) comprised 13.33% and 6.67% of the total, with no four, five, or six nucleotides being present in SSR (**Table 1**).

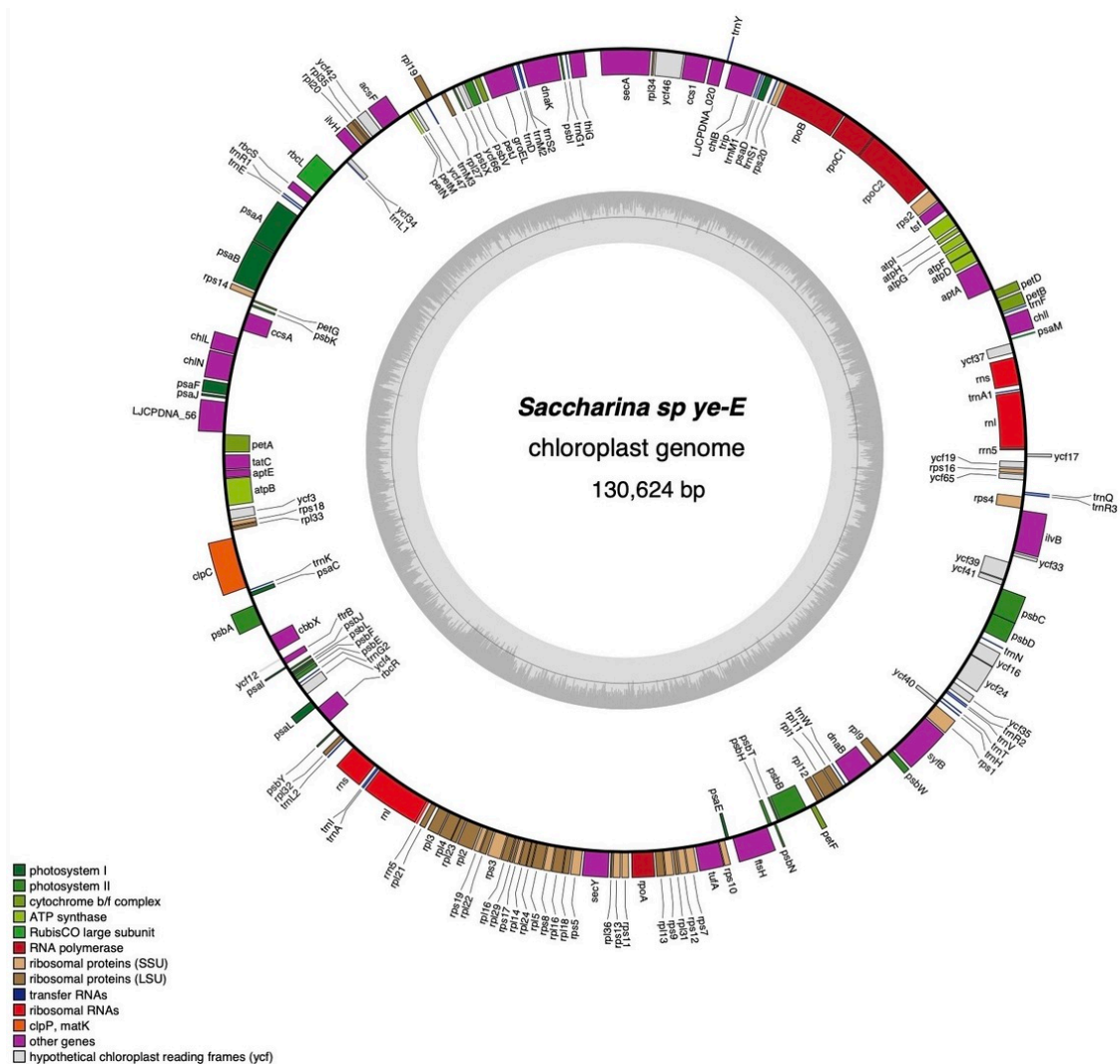


Figure 1. The outer circle represents the genome of *Saccharina sp. ye-E*.

The inner circle represents GC content. The genotypes represented by different colors are shown on the lower left.

Table 1. SSR in the chloroplast genome of *Saccharina sp. ye-E*.

Type	Repeat	Length	Number	Total
Repeats of Mononucleotide	T	10/13	5/1	6
	A	10	6	6
Repeats of Dinucleotide	TA	16	1	1
	AT	16	1	1
Repeats of Trinucleotide	ATA	15	1	1
Total			15	15

mVISTA analysis was used to compare the chloroplast genomes of *Saccharina sp. ye-E* with those of eight related species and two distant species (Outgroup), and it was observed that the related species shared the same structural composition, as shown in [Figure 1](#), and [Figure 2](#). The chloroplast genome of *Saccharina sp. ye-E* was found to be the largest and *Macrocystis integrifolia* had the smallest genome size, with a difference of around 1,300bp. The genetic composition and sequence of these chloroplast genomes were found to be highly conserved, except the *Ishige okamurae* and *Scytosiphon lomentaria*.

The use of codon bias can be employed to assess the compromise between codon divergence and natural selection in the process of translation. The EMBOSS Explorer analysis revealed that *Saccharina sp. ye-E*'s chloroplast genome had an effective codon number (ENC) of 51.666 and a codon adaptation index (CAI) of 0.617, which suggested that there was a powerful codon inclination in the chloroplast ([Figure 3](#)).

The phylogenetic analysis showed that *Saccharina sp. ye-E* was resolved in a clade with *S. latissimi* ([Figure 4](#)). Support values based on Neighbor-Joining/Mrbayes/Maximum likelihood methods were strong enough to maintained a sister relationship in *Saccharina* genus.

DISCUSSION

The characteristics of the chloroplast genome of *Saccharina sp. ye-E* were consistent with the typical features of angiosperm chloroplast genomes. SSR, or microsatellite, is defined as short sequence repeats of 1-6 bases which are found in the chloroplast genome of most plants. The amount of SSR in the chloroplast genome is lower than that in mitochondrial genomes, however it is inherited by one parent in the chloroplast genome, so it can still be used for species identification and population genetics studies.¹⁶ The presence of SSR in the chloroplast genome is essential for its molecular identification and resource conservation. Dissimilar to other reported chloroplast genomes, no four-nucleotide, five-nucleotide, or six-nucleotide SSRs were identified in the chloroplast genome of *Saccharina sp. ye-*

E. Among all genes, rRNA was found to be the most conserved, which is consistent with previous research.¹⁷ Furthermore, codon preference is affected by genome size, gene length, gene expression level, and gene density. To guarantee the accuracy and dependability of the analysis results, three methods (Neighbor-Joining, Mrbayes, Maximum likelihood) were chosen to construct the phylogenetic tree. These results concur with the phylogeny obtained from the mitochondrial genome.¹⁸ The identification of the chloroplast genome of *Saccharina sp. ye-E* will be beneficial for future algal genetics evolution and systematic studies in the Laminariaceae.

ACKNOWLEDGMENTS

This work was financially supported by Laoshan Laboratory (LSKJ202203204); the Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (grant number 2021QNLM050103-1); the National Natural Science Foundation of China (grant number 32000404, 41606038, 41976110, 42176035).

AUTHORS' CONTRIBUTION

All authors contributed to the study's conception and design. Wei Zhang and Ziwen Liu performed material preparation, data collection, and analysis. Wei Zhang wrote the first draft of the manuscript, and all authors commented on previous versions. All authors read and approved the final manuscript.

Conceptualization: Xiao Fan, Wei Zhang, Ziwen Liu;
Methodology: Wei Zhang, Ziwen Liu;
Writing - original draft preparation: Wei Zhang,
Writing – review, and editing: Xiao Fan, Wei Zhang, Ziwen Liu;
Resources: Xiao Fan.

Submitted: May 04, 2023 CST. Accepted: May 17, 2023 CST.
Published: June 07, 2023 CST.

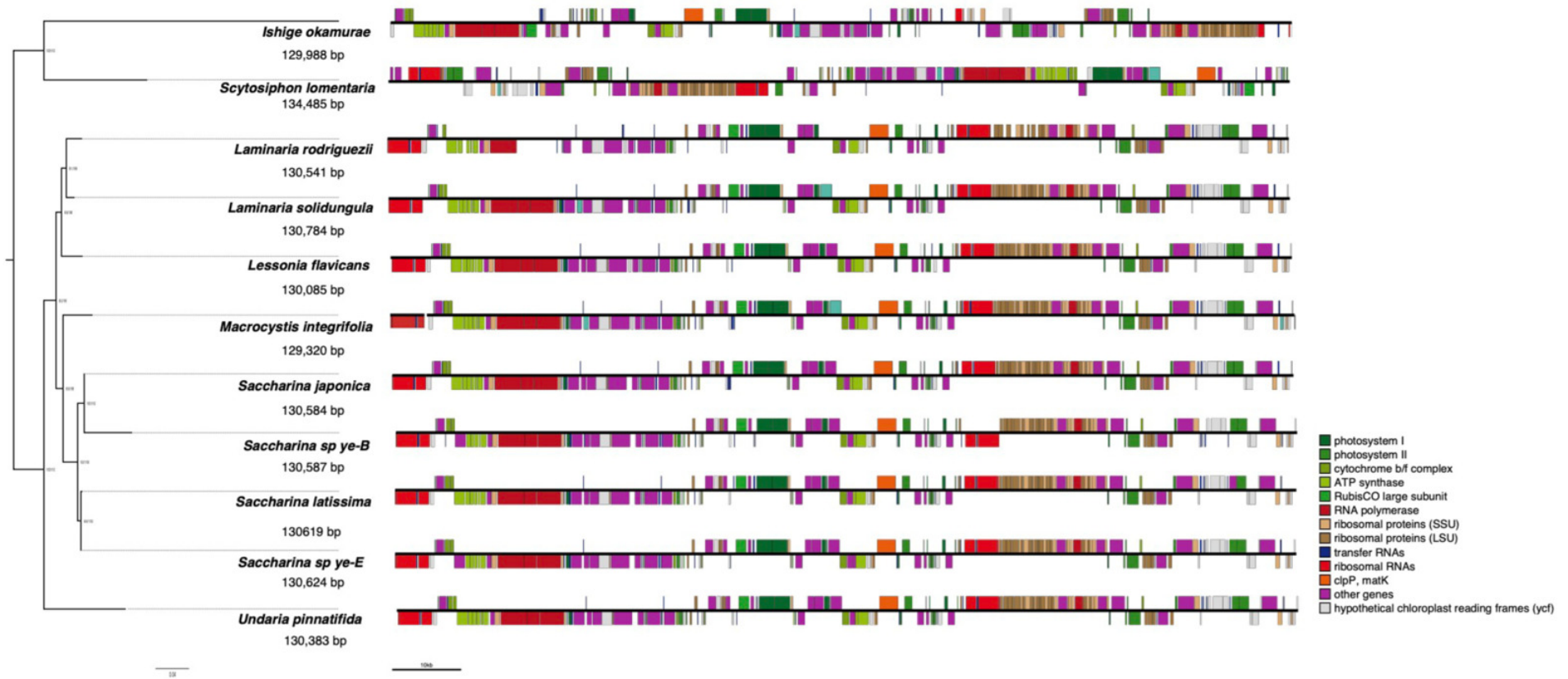


Figure 2. The phylogenetic tree and the length of chloroplast genome sequence.

Numbers in the nodes are support values of Neighbor-Joining, MrBayes test and the ML bootstrap from 1000 replicates. The genotypes represented by the colored sections are consistent with those in Figure 1.

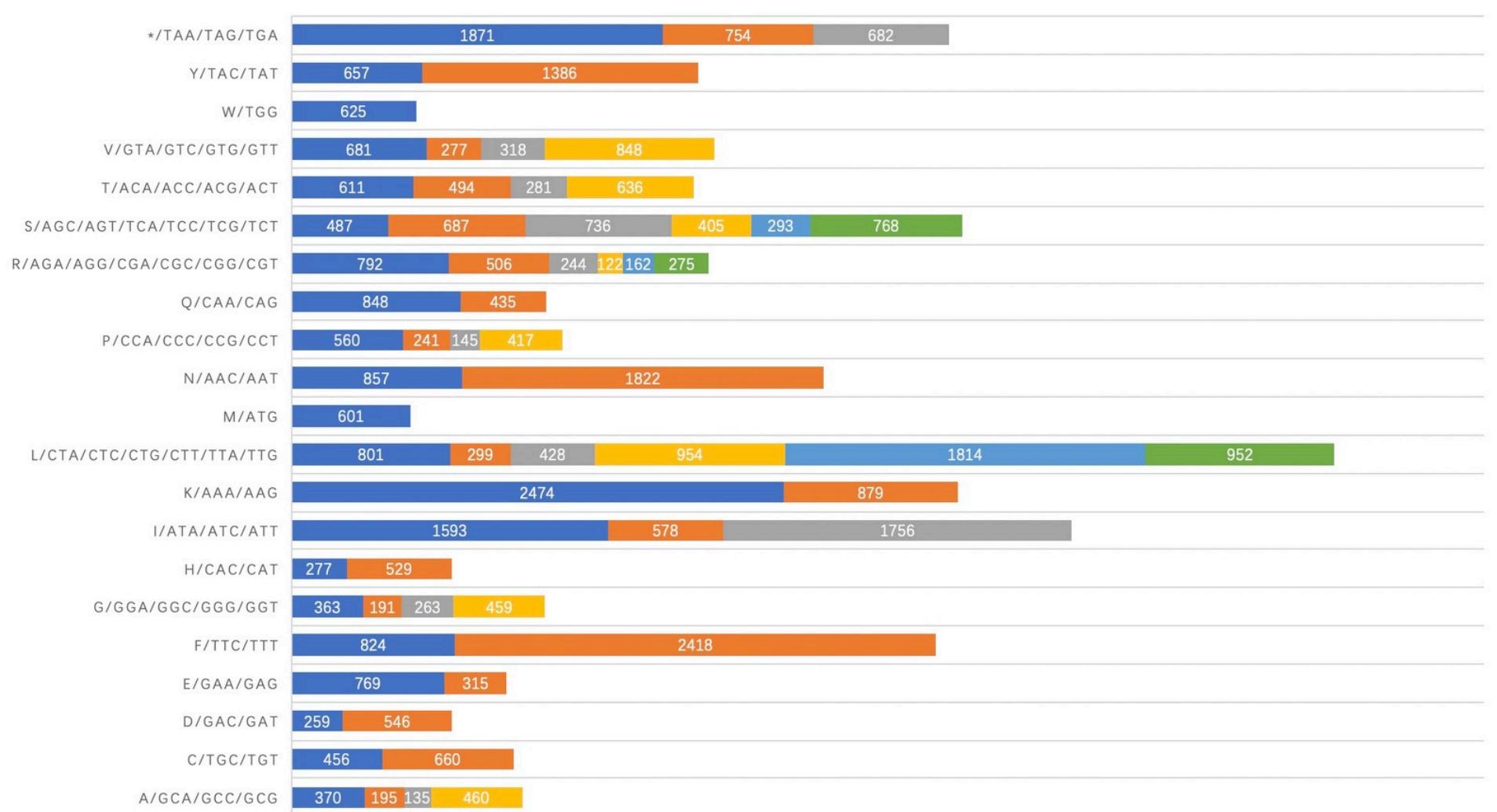


Figure 3. Distribution of relative synonymous codon usage (RSCU) in the chloroplast genome of *Saccharina sp. ye-E*.

The first letter represents the amino acid type and the rest represents the codon. *, Y, W, V, T, S, R, Q, P, N, M, L, K, I, H, G, F, E, D, C, A stand for stop codon, tyrosine, tryptophan, valine, threonine, serine, arginine, glutamine, proline, asparagine, methionine, leucine, lysine, isoleucine, histidine, glycine, Phenylalanine, glutamic acid, aspartic acid, cysteine, alanine.

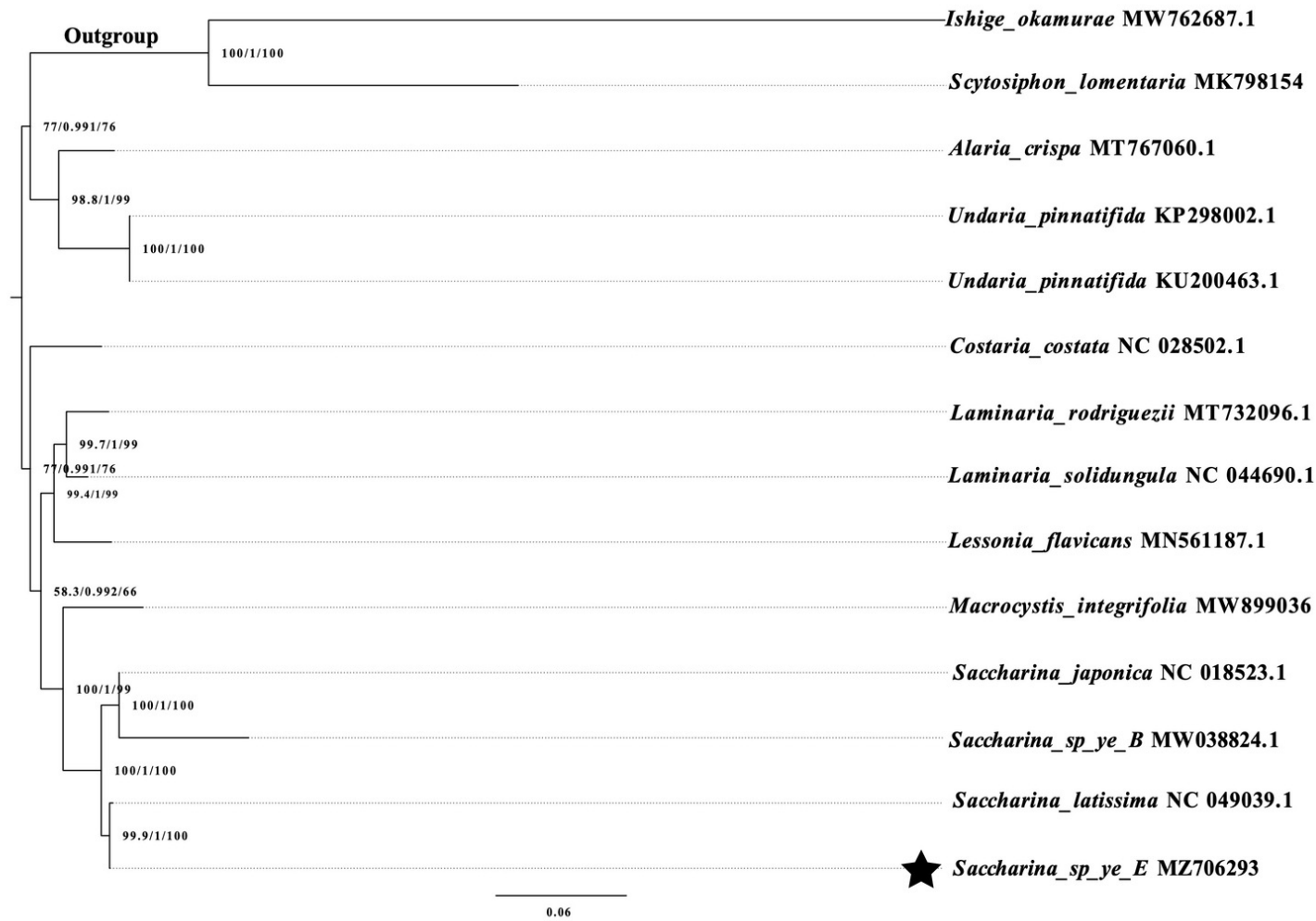


Figure 4. Neighbor-Joining/MrBayes/Maximum likelihood phylogenetic tree of *Saccharina sp. ye-E* with other species in the blown algae based on concatenated protein sequences.

Numbers in the nodes are support values of Neighbor-Joining, MrBayes test and the ML bootstrap from 1000 replicates. *Scytosiphon lomentaria* and *Ishige okamurae* were set as the outgroup. The pentagram stands for the new sequenced species in our work.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

REFERENCES

1. Hwang EK, Ha DS, Park CS. Strain selection and initiation timing influence the cultivation period of *Saccharina japonica* and their impact on the abalone feed industry in Korea. *J Appl Phycol*. 2017;29(5):2297-2305. doi:[10.1007/s10811-017-1179-2](https://doi.org/10.1007/s10811-017-1179-2)
2. Kim MK, Kim IH, Nam TJ. *Saccharina japonica* Extract Protects against Carbon Tetrachloride-induced Hepatotoxicity in Rats. *Korean Journal of Fisheries and Aquatic Sciences*. 2014;47(3):204-210. doi:[10.5657/KFAS.2014.0204](https://doi.org/10.5657/KFAS.2014.0204)
3. Vranken S, Wernberg T, Scheben A, et al. Genotype–Environment mismatch of kelp forests under climate change. *Mol Ecol*. 2021;30(15):3730-3746. doi:[10.1111/mec.15993](https://doi.org/10.1111/mec.15993)
4. Krumhansl KA, Okamoto DK, Rassweiler A, et al. Global patterns of kelp forest change over the past half-century. *Proc Natl Acad Sci USA*. 2016;113(48):13785-13790. doi:[10.1073/pnas.1606102113](https://doi.org/10.1073/pnas.1606102113). PMID:27849580
5. Satjarak A, Graham LE. Comparative DNA sequence analyses of *Pyramimonas parkeae* (Prasinophyceae) chloroplast genomes. *J Phycol*. 2017;53(2):415-424. doi:[10.1111/jpy.12515](https://doi.org/10.1111/jpy.12515)
6. Fan X, Xie W, Wang Y, Xu D, Zhang X, Ye N. The complete chloroplast genome of *Saccharina latissima*. *Mitochondrial DNA B Resour*. 2020;5(3):3481-3482. doi:[10.1080/23802359.2020.1825999](https://doi.org/10.1080/23802359.2020.1825999). PMID:33458211
7. Jackman SD, Vandervalk BP, Mohamadi H, et al. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res*. 2017;27(5):768-777. doi:[10.1101/gr.214346.116](https://doi.org/10.1101/gr.214346.116). PMID:28232478
8. Tillich M, Lehwark P, Pellizzer T, et al. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45(W1):W6-W11. doi:[10.1093/nar/gkx391](https://doi.org/10.1093/nar/gkx391). PMID:28486635
9. Jung J, Kim JI, Jeong YS, Yi G. AGORA: organellar genome annotation from the amino acid and nucleotide references. *Bioinformatics*. 2018;34(15):2661-2663. doi:[10.1093/bioinformatics/bty196](https://doi.org/10.1093/bioinformatics/bty196)
10. Shi L, Chen H, Jiang M, et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Research*. 2019;47(W1):W65-W73. doi:[10.1093/nar/gkz345](https://doi.org/10.1093/nar/gkz345). PMID:31066451
11. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res*. 2019;47(W1):W59-W64. doi:[10.1093/nar/gkz238](https://doi.org/10.1093/nar/gkz238). PMID:30949694
12. Bernt M, Donath A, Jühling F, et al. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol*. 2013;69(2):313-319. doi:[10.1016/j.ympev.2012.08.023](https://doi.org/10.1016/j.ympev.2012.08.023)
13. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res*. 2016;44(W1):W54-W57. doi:[10.1093/nar/gkw413](https://doi.org/10.1093/nar/gkw413). PMID:27174935
14. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 2019;15(1):50. doi:[10.1186/s13007-019-0435-7](https://doi.org/10.1186/s13007-019-0435-7). PMID:31139240
15. Minh BQ, Schmidt H, Chernomor O, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Published online November 21, 2019. doi:[10.1101/849372](https://doi.org/10.1101/849372)
16. Zalapa JE, Cuevas H, Zhu H, et al. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am J Bot*. 2012;99(2):193-208. doi:[10.3732/ajb.1100394](https://doi.org/10.3732/ajb.1100394)
17. Wang L, Yu X, Xu W, Zhang J, Lin H, Zhao Y. Complete chloroplast genome sequencing support *Angelica decursiva* is an independent species from *Peucedanum praeruptorum*. *Physiol Mol Biol Plants*. 2021;27(11):2503-2515. doi:[10.1007/s12298-021-01097-w](https://doi.org/10.1007/s12298-021-01097-w). PMID:34924707
18. Fan X, Wang S, Xu L, Xu D, Zhang X, Ye N. Sequencing of complete mitochondrial genome of brown algal *Saccharina* sp. ye-C5. *Mitochondrial DNA B Resour*. 2016;1(1):14-15. doi:[10.1080/23802359.2015.1137799](https://doi.org/10.1080/23802359.2015.1137799). PMID:33473390