

# Providing a phylogenetic framework for trait-based analyses in brown algae: Phylogenomic tree inferred from 32 nuclear protein-coding sequences

Shingo Akita, Christophe Vieira, Takeaki Hanyuda, Florence Rousseau, Corinne Cruaud, Arnaud Couloux, Svenja Heesch, Hiroshi Kawai, J. Mark Cock

▶ **To cite this version:**

Shingo Akita, Christophe Vieira, Takeaki Hanyuda, Florence Rousseau, Corinne Cruaud, et al.. Providing a phylogenetic framework for trait-based analyses in brown algae: Phylogenomic tree inferred from 32 nuclear protein-coding sequences. Molecular Phylogenetics and Evolution, 2022, 168, pp.107408. 10.1016/j.ympev.2022.107408 . hal-03553061

**HAL Id: hal-03553061**
**https://cnrs.hal.science/hal-03553061**

Submitted on 2 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Title:** Providing a phylogenetic framework for trait-based analyses in brown algae: phylogenomic tree inferred from 32 nuclear protein-coding sequences

Shingo Akita[1,2]*, Christophe Vieira[1]*†, Takeaki Hanyuda[1], Florence Rousseau[3], Corinne Cruaud[4], Arnaud Couloux[4], Svenja Heesch[5,6], J. Mark Cock[5], Hiroshi Kawai[1]

[1] Kobe University Research Center for Inland Seas, Rokkodai 1-1, Kobe 657-8501, Japan

[2] Ochanomizu University, Department of Biology, Ohtsuka 2-1-1, Bunkyo-ku, Tokyo 112-8610, Japan

[3] Institut de Systématique, Évolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, 57 rue Cuvier, CP 39 75005 Paris, France

[4] Génoscope, Centre National de Séquençage, 2 rue G. Crémieux, Evry, CP 5706, France

[5] CNRS, Sorbonne Université, Algal Genetics Group, UMR 8227 Integrative Biology of Marine Models, Station Biologique, F 29688, Roscoff, France

[6] Universität Rostock Institut für Biowissenschaften Angewandte Ökologie & Phykologie Albert-Einstein-Straße 21 18059 Rostock

* Shared co-first authorship.

† corresponding author: cvcarp@gmail.com

**Abbreviations**

**BACR:** brown algal crown radiation

**SSDO:** Sphacelariales, Syringodermatales, Dictyotales, Onslowiales

Journal Pre-proofs

**ABSTRACT**

In the study of the evolution of biological complexity, a reliable phylogenetic framework is needed. Many attempts have been made to resolve phylogenetic relationships between higher groups (i.e., interordinal) of brown algae (Phaeophyceae) based on molecular evidence, but most of these relationships remain unclear. Analyses based on small multi-gene data (including chloroplast, mitochondrial and nuclear sequences) have yielded inconclusive and sometimes contradictory results. To address this problem, we have analyzed 32 nuclear protein-coding sequences in 39 Phaeophycean species belonging to eight orders. The resulting nuclear-based phylogenomic trees provide virtually full support for the phylogenetic relationships within the studied taxa, with few exceptions. The relationships largely confirm phylogenetic trees based on nuclear, chloroplast and mitochondrial sequences, except for the placement of the Sphacelariales with weak bootstrap support. Our study indicates that nuclear protein-coding sequences provide significant support to conclusively resolve phylogenetic relationships among Phaeophyceae, and may be a powerful approach to fully resolve interordinal relationships with increased taxon sampling.

**Keywords:** Ectocarpales; Genomics; Heterokontophytes; Phaeoexplorer; Phaeophyceae; Stramenopiles

Journal Pre-proofs

## 1. Introduction

Brown algae or Phaeophyceae represent a large class of multicellular photosynthetic marine eukaryotes distributed in cold to tropical marine waters with major ecological and economical functions (Bringloe et al., 2020). Brown algae include about 2000 species classified into 340 genera, 59 families and 19 orders (Guiry and Guiry, 2021). In contrast to land plants, our understanding of brown algal evolution remains quite limited (Bringloe et al., 2020). The Phaeoexplorer project (https://phaeoexplorer.sb-roscoff.fr/home/) was initiated in 2016 with the objective of exploring the evolution of biological complexity in brown algae by generating genomic data for a broad range of brown algal species. A fully resolved backbone phylogeny for brown algae is required to investigate questions in evolution, ecology and conservation.

Initial efforts to elucidate phylogenetic relationships among brown algae were based on nuclear ribosomal DNA sequences (Bhattacharya et al., 1992; Druehl et al., 1997; Rousseau and De Reviers, 1999; Tan and Druehl, 1993), followed by studies adding RuBisCO genes (Bailey and Andersen, 1999; De Clerck et al., 2006; Draisma et al., 2001; Peters and Ramírez, 2001; Phillips et al., 2008). More recently, significant attempts were made to produce multi-gene datasets but with a limited number of genes (less than 15), mainly from chloroplast and mitochondrial genomes with nuclear DNA still represented only by ribosomal markers (Bringloe et al., 2020; Cho et al., 2012; Kawai et al., 2015; Lane et al., 2007; Silberfeld et al., 2010). While contributing considerably to our understanding of brown algal evolutionary history, these previous phylogenies only partially resolved relationships at some taxonomic levels, in particular for interordinal relationships within the Fucophycideae. Furthermore, in

Journal Pre-proofs

some cases, multi-gene molecular phylogenies based on chloroplast and mitochondrial genes indicated contradictory relationships (Bringloe et al., 2020; Kawai et al., 2015; Kawai et al., 2017; Kawai et al., 2013). Whole genome data are gradually emerging for brown algae (Bringloe et al., 2020), and several studies based on mitochondrial or chloroplast genomes, focused mainly on Laminariales and Fucales members, produced strong evidence supporting the merit of genome data in resolving phylogenetic relationships (e.g., Zhang et al., 2019). However, so far, whole genome data are available for only six of the 19 orders of brown algae (Bringloe et al., 2020). More recently, two studies made use of nuclear transcriptomic data to deduce brown algal evolutionary relationships (Jackson et al., 2017; Sun et al., 2014). The latter study, which was based on 108 orthologous genes from 18 Phaeophyceae species belonging to six orders (i.e., Desmarestiales, Dictyotales, Ectocarpales, Fucales, Ishigeales, Laminariales), produced mixed results (Sun et al., 2014). Overall, phylogenomic trees provided strong to full support for infraordinal relationships and for the relationship between Ectocarpales and Laminariales, but less strong support for other interordinal relationships (e.g., bootstrap (BS) = 55 for Dictyotales-Ishigeales node) (Sun et al., 2014). A second nuclear-transcriptome-based phylogenomic study conducted at the infraordinal level, employed 152 genes from 15 Laminariales species and yielded a nearly fully resolved tree, which enabled relationships between various genera to be revised (Jackson et al., 2017). Family-level relationships were similar to those previously established based on a dataset of eight chloroplast and mitochondrial genes (Kawai et al., 2013). These two transcriptome-based studies effectively resolved lower taxonomic-level phylogenetic relationships (infraordinal), but further studies are needed to improve resolving power at a higher level (i.e., interordinal level).

Journal Pre-proofs

The objective of the study presented here was twofold: (1) to test the power of nuclear genome-based datasets for resolving inter and infraordinal phylogenetic relationships in brown algae and (2) to provide a reliable phylogenetic framework for trait-based phylogenetic analyses in the context of the Phaeoexplorer project. Maximum likelihood and Bayesian phylogenomic trees were constructed based on the predicted amino-acid sequence data of 32 nuclear protein-coding genes in 39 brown algal species and two outgroup taxa selected in the Phaeoexplorer project.

## 2. Materials and methods

### *2.1. Taxa sampling*

A total of 39 species of brown algae representing 24 genera, 16 families and eight orders (i.e., Desmarestiales, Discosporangiales, Dictyotales, Ectocarpales s.l., Fucales, Laminariales, Sphacelariales, Tilopteridales), in addition to two outgroup taxa, *Schizocladia ischiensis* E.C.Henry, K.Okuda & H.Kawai (class Schizocladiophyceae) and *Phaeothamnion wetherbeei* R.A.Andersen, L.Graf & H.S.Yoon (class Phaeothamniophyceae), selected in the Phaeoexplorer project, were used in the present study for the phylogenetic reconstructions (Table S1).

### *2.2. Culture conditions, DNA extraction, genome sequencing and assembly*

Genomic DNA was extracted from laboratory-housed culture strains (Roscoff Culture Collection, Bezhin Rosko Culture Collection, Kobe University Macroalgal Culture Collection, Culture Collection of Algae at Göttingen University) or from specimens collected in the field (Table S1). For species with diplohaplontic life cycles, the gametophyte generation was grown in culture either in 140 mm Petri dishes or in 2-10 L bottles, the latter aerated by bubbling with sterile air. Standard growth conditions

were sterile, Provosoli-enriched (Starr and Zeikus, 1993) natural seawater (ES medium) under fluorescent white light (10-30 μM photons/$m^2$·s) at 13°C. The freshwater species *Pleurocladia lacustris* A.Braun and *Heribaudiella fluviatilis* (Areschoug) Svedelius were grown in natural seawater that had been diluted to 5% with distilled water (i.e., 95% distilled water / 5% seawater) before addition of ES medium (http://sagdb.uni-goettingen.de/culture_media/01%20Basal%20Medium.pdf) micronutrients, at 20°C for *P. lacustris*. *Phaeothamnion wetherbeei* was grown in MIEB12 (medium 7 in (Schlösser, 1994)). Material for fucoid species was collected on the seashore, and extractions used either dissected meristematic regions or released male gametes. DNA was extracted with either the Nucleospin Plant II midi DNA Extraction Kit (Macherey-Nagel, Düren, Germany) or the OmniPrep Genomic DNA Purification Kit (G Biosciences, St. Louis, MO, USA). DNA quality was assessed using a Qubit fluorometer (Themo Fisher Scientific, Waltham, MA, USA), and fragment length was assessed by migration on a 1% agarose gel for some of the samples. DNA (30–100 ng) was sonicated to a 100 to 800 bp size range using a Covaris E220 sonicator (Covaris, Woburn, MA, USA). Fragments were end-repaired, 3′-adenylated and Illumina adapters (Bioo Scientific, Austin, TX, USA) were then added using the NEBNext Sample Reagent Set (New England Biolabs, Ipswich, MA, USA). Ligation products were purified using Ampure XP (Beckmann Coulter Genomics, Danvers, MA, USA), and DNA fragments (>200 bp) were PCR amplified using Illumina adapter-specific primers and the KAPA HiFi HotStart polymerase (KapaBiosystems, Wilmington, MA, USA). Libraries were quantified by qPCR using the KAPA Library Quantification Kit for Illumina Libraries (KapaBiosystems), and library profiles were assessed using a DNA High Sensitivity LabChip kit on an Agilent

Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Some samples were prepared according to the same protocol but with less DNA input (10-25ng) and using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs). Illumina short read sequence data (150 bp paired end sequence) was generated on either a HiSeq4000 or a NovaSeq6000 platform. The sequence data was first assembled with Megahit (Li et al., 2015), and eukaryotic contigs were separated from the bacterial contigs using a gene detection strategy based on Metagen and Blastp. The bacterial contigs were then used to filter the initial dataset by mapping Illumina reads to the contigs with BWA and then removing these reads, which were not of eukaryotic origin. The set of filtered Illumina reads were then assembled with SPAdes (Bankevich et al., 2012) using the standard protocol.

### *2.3. Identification of orthologous single-copy genes*

Genomes were predicted from *Schizocladia ischiensis* and 23 Phaeophycean species that belong to various orders (Table S1). Predicted genomes were analyzed with OrthoFinder v.2.3.11 (Emms and Kelly, 2015) to detect orthologous single-copy genes. The program detected a total of 32 single-copy genes, which were used in the phylogenetic analyses. Table S2 presents a list of the 32 orthologues in the reference *Ectocarpus* sp.7 strain Ec32 genome. For *Phaeothamnion wetherbeei* and the other 16 species, orthologous amino acid sequences were retrieved using the "tblastn" function of BLAST+ 2.6.0 (Camacho et al., 2009).

### *2.4. Phylogenetic analyses*

Amino acid sequences of the thirty-two single-copy genes were aligned manually using the alignment software AliView v.1.26 (Larsson, 2014). The aligned sequences were concatenated (41 OTUs, 15,597 amino acids (aa)) and subjected to

Maximum likelihood (ML) and Bayesian (BI) analysis. For ML analysis, 10,000 Rapid Bootstrap searches and the subsequent ML search with gamma model were conducted using RAxML v.8.2.9 (Stamatakis, 2014). Bayesian analyses were performed using MrBayes v.3.2.2 (Ronquist et al., 2012). The best-fit evolutionary models for each gene were determined on the basis of AIC for ML and BIC for BI using Kakusan4 (Tanabe, 2011) (Table S2). The Bayesian analyses were initiated with a random starting tree, and four chains of Markov chain Monte Carlo iterations were run simultaneously for 10,000,000 generations, keeping one tree every 100 generations. The first 10,000 trees sampled were discarded as 'burn-in', based on the stationarity of log-likelihood as assessed using Tracer v.1.7.1 (Rambaut et al., 2018). A consensus topology and posterior probability values were calculated from the remaining trees.

## 3. Results

### *3.1. Data Characteristics*

We assembled a 41-species dataset based on 32 nuclear genes (Tables S2, S3) from 39 species of Phaeophyceae, belonging to 24 genera, 16 families and eight orders (Desmarestiales, Discosporangiales, Dictyotales, Ectocarpales, Fucales, Lamiariales, Sphacelariales, Tilopteridales), together with one species from the Schizocladiophyceae (*Schizocladia ischiensis*) and one species from the Phaeothamniophyceae (*Phaeothamnion wetherbeei*), which served as outgroups (Table S2). The gene matrix was 95.6% complete and the dataset contained 15,597 aa (Table S2).

### *3.2. Phylogenetic relationships*

ML and BI analyses of the 32-nuclear concatenated genes in the 41-species dataset

Journal Pre-proofs

produced trees with identical topology except for the placement of the Sphacelariales in relation to the Dictyotales (Figs 1, S1). In both analyses, placement of the Sphacelariales was poorly supported with a BS of 53 % and a Bayesian posterior probability (BPP) of 59 %. Eighty-two percent of the 38 brown algal nodes were fully supported in the ML analysis (i.e., 100 % BS) and 95 % in BI analysis (i.e., 100 % BPP). Five nodes within the genus *Ectocarpus* yielded BS values of 81 to 99 %. The seven inter-family relationships presented 99 – 100 % BS and BPP support, and the seven interordinal relationships were fully supported for both BS and ML analyses with the exception of the Sphacelariales. *Porterinema fluviatile* (H.C.Porter) Waern (SAG 2381) positioned among the Ectocarpales as a sister to the Chordariaceae.

**4. Discussion**

Morphology and gene-based studies of higher-level brown algal phylogeny have a long history (De Reviers et al., 2007; Draisma et al., 2001; Silberfeld et al., 2010; Sun et al., 2014). However, many higher-level relationships remain problematic, even with the major improvements that have resulted from multi-gene methodologies (Bringloe et al., 2020; Silberfeld et al., 2010). Recent studies have demonstrated the value of studies based on multi-gene data from organellar (Cui et al., 2019; Liu and Pang, 2015a, b; Liu et al., 2019; Liu et al., 2020; Wu et al., 2021; Zhang et al., 2019) and nuclear genomes (Sun et al., 2014) for solving higher-level relationships in the Phaeophyceae. The current study was based on higher ordinal coverage of protein-coding nuclear genes to assess the resolving power of protein-coding nuclear genes in brown algae at this taxonomic level.

The approach used for the current study significantly improves on previous

analyses. With the exception of Sphacelariales, interordinal relationships were fully resolved in the current study. Sun et al. (2014) used a similar approach to resolve interordinal relationships, but it is difficult to compare our results with this earlier study, as it was based on different nuclear genes and taxon sampling. However, some conclusions can still be drawn. First, our study included a smaller number of nuclear genes (32 vs. 108 genes), but comprised a higher number of taxa (39 vs. 18 species, 24 vs. 10 genera, 16 vs. seven families, and eight vs. six orders) than that of Sun et al. (2014). The strong branch support obtained by our phylogenetic analyses indicates that the nuclear genes selected in this study are sufficient to generate a robust molecular higher-level phylogeny, although the number of genes are lower than that used by Sun et al. (2014). Testing the quantity of sequence data required for strong node support (e.g., Wortley et al., 2005) will facilitate targeted inclusion of additional taxa in future studies. Second, a better taxonomic representativity of the different orders seems to increase node robustness in our study. Five of the orders analyzed by Sun et al. (2014) were also analyzed in our study, but we did not include any members of the Ishigeales. Another important difference with Sun et al. (2014) concerns the choice of outgroup. In their study, they used a distant outgroup, i.e., the diatom *Phaeodactylum tricornutum* (Bacillariophyceae), which caused a very long branch and may have affected the ingroup taxa relationships and caused lower bootstrap supports (Li et al., 2014), while we used closely related outgroups, i.e., Scizocladiophyceae and Phaeothamniophyceae, which is recommended, when distantly related ingroup taxa are poorly sampled (Li et al., 2014), as it is the case in Sun et al. (2014) and our studies. The taxa analyzed by Sun et al. (2014) were biased towards Fucales and Ectocarpales, and our study mainly towards Ectocarpales. The Sun et al. (2014) analysis included four brown algal crown radiation (BACR) orders and our study five. In line with

previous multi-gene studies, both studies conclusively resolved the sister relationship between the Ectocarpales and the Laminariales. The Sun et al. (2014) analysis did not fully resolve relationships among the BACR orders (Desmarestiales, Ectocarpales, Fucales, Laminariales), but interordinal relationships were fully resolved in our analyses (Desmarestiales, Ectocarpales, Fucales, Laminariales, Tilopteridales). Our analyses conclusively positioned the Desmarestiales as first divergence within the BACR with almost full support (BS = 99, BPP = 100). This stands in contrast with previous nuclear- and mitochondrial-based analyses (Liu and Pang, 2015a, b; Liu et al., 2019; Sun et al., 2014), which positioned the Fucales basally in relation to the Desmarestiales. Nevertheless, in the mitochondrial-based analyses (Liu and Pang, 2015a, b; Liu et al., 2019), the ordinal taxon coverage was very limited, and in the nuclear-based analysis (Sun et al., 2014), placement of the Desmarestiales and the Fucales was supported by lower BS values of 95 and 90, respectively. These low support values in Sun et al. (2014) may be explained by the long branch of the outgroup. Overall, these results indicate a beneficial effect of increased taxon sampling for phylogenetic inference in brown algae using protein-coding nuclear data. Since our dataset only encompassed eight of the 19 Phaeophyceae orders, we assume that increasing taxon sampling may further resolve the placement of Sphacelariales.

Interordinal relationships retrieved in this study were nearly identical to those established by Bringloe et al. (2020) based on 12 plastid, mitochondrial and nuclear markers (18S, 5.8S, 28S, *atp*B, *psb*A, *psa*B, *psa*A, *rbc*L, *psb*C, *cox*1, *cox*3, *nad*1), with the exception of the relationship between the Dictyotales and the Sphacelariales. The phylogenetic position of the Sphacelariales, associated to the monophyly of the Dictyotophycidae, was found to be unstable in previous studies. The weak support and unstable position of the Sphacelariales in most studies is possibly

associated with incongruence among markers, with a weak phylogenetic signal and/or methodological bias. We assume that adding the missing orders for the Sphacelariales, Syringodermatales, Dictyotales, Onslowiales (SSDO) clade (i.e., Onslowiales and Syringodermatales) may help improving the resolution.

Our dataset included five out of the seven families of Ectocarpales, and relationships within Ectocarpales *sensu lato* were fully resolved. The Acinetosporaceae were positioned as a sister group to the Chordariaceae. Familial relationships within the Laminariales were fully supported and identical to the most up-to-date phylogeny of this order based on genomic data (Starko et al., 2019). For *Porterinema fluviatile,* it should be noted that strain SAG 2381 differs genetically from strain SAG 124.79 (McCauley and Wehr, 2007), which are both from the Culture Collection of Algae at Göttingen University (SAG). SAG 2381 was positioned within the Ectocarpales in this study, while SAG 124.79 was positioned basally among the brown algae (Draisma et al., 2010; McCauley and Wehr, 2007). Molecular and morphological analyses of the culture strains of SAG 2381, SAG 124.79, and *Pilinia rimosa* collected from various localities worldwide indicated that SAG 2381 is the true *P. fluviatile* displaying the characteristic plurilocular zoidangia, whereas SAG 124.79 is in fact *P. rimosa* (Kawai et al., 2021).

**5. Conclusions**

In conclusion, this study provides a valuable phylogenetic framework for the Phaeoexplorer project, which focuses on the study of the evolution of biological complexity in brown algae, and further evidence supporting the resolving power of protein-coding nuclear genes at inter-ordinal level in brown algae, demonstrating the interest of nuclear multi-gene analysis as a complementary tool to analysis of

Journal Pre-proofs

mitochondrial and chloroplast organellar genome data. Since the current study only included eight of the 19 orders of brown algae, much effort is still needed to provide a comprehensive phylogeny robustly resolving inter-ordinal relationships in brown algae. By comparing our results with previous studies, taxonomic sampling appears to be a major element in resolving interordinal relationships. In the data vs. taxa trade-off, it would therefore be better to give priority to taxonomic sampling. Given the satisfactory results based on the 32 nuclear genes included in the present study, we recommend that future studies make use of the same set of genes as a standard approach using whole-genome sequencing data.

**Acknowledgements**

We would like to thank the following colleagues for contributions to data production and treatment: Akira Peters (Behzin Rosco), Zofia Nehr, Olivier Godfroy, Delphine Scornet, Susana Coelho, Loraine Guéguen, Erwan Corre, Chloé Jollivet (Station Biologique de Roscoff), France Denoeud (Genoscope). We are grateful to Line Le Gall for her helpful comments on the manuscript. The authors thank the Japan Society for the Promotion of Science (JSPS) for financial support to SA and CV.

**Funding**

This work was supported by the France Génomique National infrastructure project Phaeoexplorer (ANR-10-INBS-09), the ERC grant SEXSEA (638240) and the JSPS Grant-in-Aid for Scientific Research (19F19796).

Journal Pre-proofs

**References**

Bailey, J.C., Andersen, R.A., 1999. Analysis of clonal cultures of the brown tide algae *Aureococcus* and *Aureoumbra* (Pelagophyceae) using 18S rRNA, *rbc*L, and rubisco spacer sequences. J. Phycol. 35, 570-574.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455-477.

Bhattacharya, D., Medlin, L., Wainright, P.O., Ariztia, E.V., Bibeau, C., Stickel, S.K., Sogin, M.L., 1992. Algae containing chlorophylls a+ c are paraphyletic: molecular evolutionary analysis of the Chromophyta. Evolution 46, 1801-1817.

Bringloe, T.T., Starko, S., Wade, R.M., Vieira, C., Kawai, H., De Clerck, O., Cock, J.M., Coelho, S.M., Destombe, C., Valero, M., Neiva, J., Pearson, G.A., Faugeron, S., Serrao, E.A., Verbruggen, H., 2020. Phylogeny and evolution of the brown algae. Crit. Rev. Plant Sci. 39, 281-321.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 1-9.

Cho, S.M., Lee, S.M., Ko, Y.D., Mattio, L., Boo, S.M., 2012. Molecular systematic reassessment of *Sargassum* (Fucales, Phaeophyceae) in Korea using four gene regions. Bot. Mar. 55, 473-484.

Cui, Y., Liu, T., Wang, X., Qu, J., Jia, X., 2019. The complete chloroplast genome of Sargassum horneri and its phylogenetic analysis. Mitochondrial DNA Part B 4, 3312-3313.

De Clerck, O., Leliaert, F., Verbruggen, H., Lane, C.E., De Paula, J.C., Payo, D.A.,

Journal Pre-proofs

Coppejans, E., 2006. A revised classification of the Dictyoteae (Dictyotales, Phaeophyceae) based on *rbc*L and 26S ribosomal DNA sequence analyses. J. Phycol. 42, 1271-1288.

De Reviers, B., Rousseau, F., Draisma, S.G.A., 2007. Classification of the Phaeophyceae from past to present and current challenges. In: Brodie, J., Lewis, J. (Eds.), Unravelling the algae: the past, present, and future of algal systematics. CRC Press, Taylor & Francis Group, , Boca Raton, FL, USA, p. 267.

Draisma, S.G., Prud'Homme van Reine, W.F., Stam, W.T., Olsen, J.L., 2001. A reassessment of phylogenetic relationships within the Phaeophyceae based on RUBISCO large subunit and ribosomal DNA sequences. J. Phycol. 37, 586-603.

Draisma, S.G., Prud'homme Van Reine, W.F., Kawai, H., 2010. A revised classification of the Sphacelariales (Phaeophyceae) inferred from a psb C and rbc L based phylogeny. Eur. J. Phycol. 45, 308-326.

Druehl, L., Mayes, C., Tan, I., Saunders, G., 1997. Molecular and morphological phylogenies of kelp and associated brown algae. Origins of Algae and their Plastids. Springer, pp. 221-235.

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome biology 16, 157.

Guiry, M.D., Guiry, G.M., 2021. AlgaeBase. World-wide electronic publication. National University of Ireland, Galway. http://www.algaebase.org (accessed 16 April 2021).

Jackson, C., Salomaki, E.D., Lane, C.E., Saunders, G.W., 2017. Kelp transcriptomes

Journal Pre-proofs

provide robust support for interfamilial relationships and revision of the little known Arthrothamnaceae (Laminariales). J. Phycol. 53, 1-6.

Kawai, H., Hannyuda, T., Henry, E.C., 2021. Transfer of *Pilinia* from Ectocarpales to Ishigeales (Phaeophyceae) with proposal of Piliniaceae fam. nov., and taxonomy of *Porterinema* in Ectocarpales. Eur. J. Phycol.

Kawai, H., Hanyuda, T., Draisma, S.G., Wilce, R.T., Andersen, R.A., 2015. Molecular phylogeny of two unusual brown algae, *Phaeostrophion irregulare* and *Platysiphon glacialis*, proposal of the Stschapoviales ord. nov. and Platysiphonaceae fam. nov., and a re-examination of divergence times for brown algal orders. J. Phycol. 51, 918-928.

Kawai, H., Hanyuda, T., Gao, X., Terauchi, M., Miyata, M., Lindstrom, S.C., Klochkova, N.G., Miller, K.A., 2017. Taxonomic revision of the Agaraceae with a description of *Neoagarum* gen. nov. and reinstatement of *Thalassiophyllum*. J. Phycol. 53, 261-270.

Kawai, H., Hanyuda, T., Ridgway, L.M., Holser, K., 2013. Ancestral reproductive structure in basal kelp *Aureophycus aleuticus*. Sci. Rep. 3, 1-7.

Lane, C.E., Lindstrom, S.C., Saunders, G.W., 2007. A molecular assessment of northeast Pacific *Alaria* species (Laminariales, Phaeophyceae) with reference to the utility of DNA barcoding. Mol. Phylogen. Evol. 44, 634-648.

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30, 3276-3278.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674-1676.

Li, T., Hua, J., Wright, A.M., Cui, Y., Xie, Q., Bu, W., Hillis, D.M., 2014.

Journal Pre-proofs

Long-branch attraction and the phylogeny of true water bugs (Hemiptera: *Nepomorpha*) as estimated from mitochondrial genomes. BMC Evol. Biol. 14, 1-12.

Liu, F., Pang, S., 2015a. Mitochondrial genome of *Turbinaria ornata* (Sargassaceae, Phaeophyceae): comparative mitogenomics of brown algae. Curr. Genet. 61, 621-631.

Liu, F., Pang, S., 2015b. Mitochondrial phylogenomics reveals a close relationship between *Petalonia fascia* (Scytosiphonaceae, Phaeophyceae) and *Ectocarpus siliculosus*. J. Appl. Phycol. 27, 1021-1028.

Liu, F., Zhang, Y., Bi, Y., Chen, W., Moejes, F.W., 2019. Understanding the evolution of mitochondrial genomes in Phaeophyceae inferred from mitogenomes of *Ishige okamurae* (Ishigeales) and *Dictyopteris divaricata* (Dictyotales). J. Mol. Evol. 87, 16-26.

Liu, T., Cui, Y., Jia, X., Chen, B., Ma, Z., Zou, H., Wang, S., Wu, M., 2020. The complete chloroplast genome of Sargassum fusiforme. Mitochondrial DNA Part B 5, 576-577.

McCauley, L.A., Wehr, J.D., 2007. Taxonomic reappraisal of the freshwater brown algae *Bodanella*, *Ectocarpus*, *Heribaudiella*, and *Pleurocladia* (Phaeophyceae) on the basis of *rbc*L sequences and morphological characters. Phycologia 46, 429-439.

Peters, A.F., Ramírez, M.E., 2001. Molecular phylogeny of small brown algae, with special reference to the systematic position of *Caepidium antarcticum* (Adenocystaceae, Ectocarpales). Cryptogamie Algol 22, 187-200.

Phillips, N., Burrowes, R., Rousseau, F., De Reviers, B., Saunders, G.W., 2008. Resolving evolutionary relationships among the brown algae using chloroplast

Journal Pre-proofs

and nuclear genes. J. Phycol. 44, 394-405.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 10.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539-542.

Rousseau, F., De Reviers, B., 1999. Phylogenetic relationships within the Fucales (Phaeophyceae) based on combined partial SSU+ LSU rDNA sequence data. Eur. J. Phycol. 34, 53-64.

Schlösser, U.G., 1994. SAG-Sammlung von Algenkulturen at the University of Göttingen Catalogue of Strains 1994. Botanica acta 107, 113-186.

Silberfeld, T., Leigh, J.W., Verbruggen, H., Cruaud, C., De Reviers, B., Rousseau, F., 2010. A multi-locus time-calibrated phylogeny of the brown algae (Heterokonta, Ochrophyta, Phaeophyceae): investigating the evolutionary nature of the "brown algal crown radiation". Mol. Phylogen. Evol. 56, 659-674.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.

Starko, S., Gomez, M.S., Darby, H., Demes, K.W., Kawai, H., Yotsukura, N., Lindstrom, S.C., Keeling, P.J., Graham, S.W., Martone, P.T., 2019. A comprehensive kelp phylogeny sheds light on the evolution of an ecosystem. Mol. Phylogen. Evol. 136, 138-150.

Starr, R.C., Zeikus, J.A., 1993. Utex—The culture collection of algae at the University of Texas at Austin 1993 List of Cultures. J. Phycol. 29, 1-106.

Sun, J., Wang, L., Wu, S., Wang, X., Xiao, J., Chi, S., Liu, C., Ren, L., Zhao, Y., Liu,

Journal Pre-proofs

T., 2014. Transcriptome-wide evolutionary analysis on essential brown algae (Phaeophyceae) in China. Acta Oceanol. Sin. 33, 13-19.

Tan, I.H., Druehl, L.D., 1993. Phylogeny of the Northeast Pacific brown algal (Phaeophycean) orders as inferred from 18S rRNA gene sequences. Fourteenth International Seaweed Symposium. Springer, pp. 699-704.

Tanabe, A.S., 2011. Kakusan4 and Aminosan: two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. Mol. Ecol. Resour. 11, 914-921.

Wortley, A.H., Rudall, P.J., Harris, D.J., Scotland, R.W., 2005. How much data are needed to resolve a difficult phylogeny? Case study in Lamiales. Syst. Biol. 54, 697-709.

Wu, X., Zhang, P., Zhang, Y., Wang, T., 2021. The complete chloroplast genome of *Sargassum hemiphyllum* var. *Chinense* (Sargassaceae, Phaeophyceae) and its phylogenetic analysis. Mitochondrial DNA Part B 6, 278-279.

Zhang, Y., Wang, S., Qian, W., Shi, Y., Li, N., Yan, X., Zou, H., Wu, M., 2019. Characterization of the complete chloroplast genome of *Sargassum fusiforme* and its phylogenomic position within phaeophyceae. Mitochondrial DNA Part B 4, 3258-3259.

Journal Pre-proofs

**Figure legend**

**Fig. 1.** Maximum likelihood phylogenomic tree based on 39 brown algal taxa and two outgroup species inferred from the amino-acid sequence data of 32 nuclear protein-coding genes. '*' indicates ML bootstrap = 100.

**Appendix A. Supplementary data**

**Table S1.** Algal specimens used in this study.

**Table S2.** List of strains and single-copy genes analyzed in this study with the coverage (%) for each species and for each of the single-copy genes.

**Figure S1.** Bayesian phylogenomic tree based on 39 brown algal taxa and two outgroup species inferred from the amino-acid sequence data of 32 nuclear protein-coding genes. '*' indicates posterior probabilities = 100.

**Shingo Akita:** Conceptualization, Formal analysis, Investigation, Methodology, Data curation, Visualization, Writing - Review & Editing; **Christophe Vieira**: Conceptualization, Formal analysis, Investigation, Methodology, Data curation, Visualization, Writing - Original Draft; **Takeaki Hanyuda**: Formal analysis, Writing - Review & Editing; **Florence Rousseau***:* Writing - Review & Editing; **Corinne Cruaud**: Formal analysis; **Arnaud Couloux:** Formal analysis; **Svenja Heesch:** Formal analysis, Writing - Review & Editing; **J. Mark Cock:** Resources, Writing - Review & Editing, Funding acquisition, Project administration; **Hiroshi Kawai:** Conceptualization, Resources, Writing - Review & Editing, Project administration, Funding acquisition

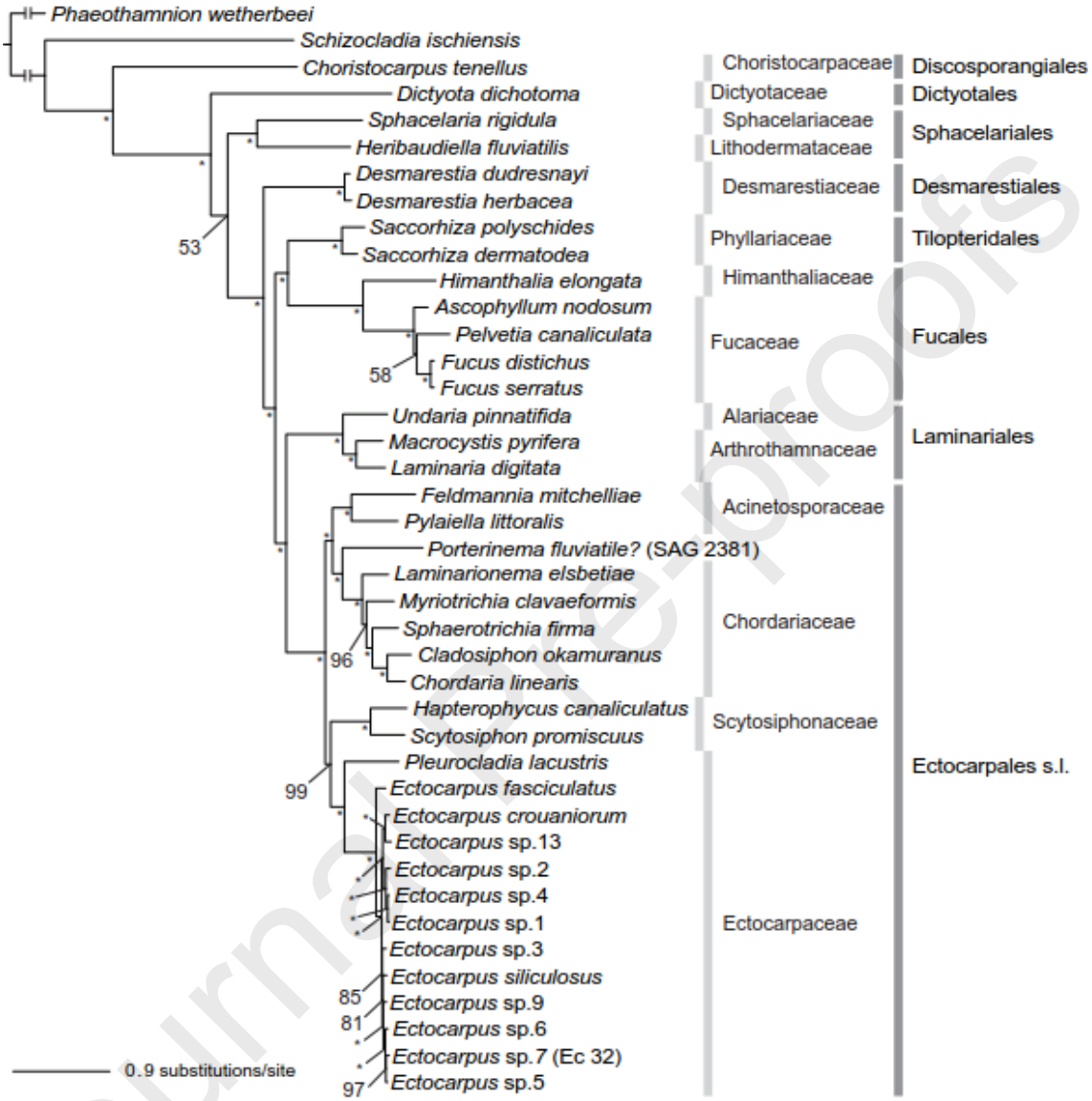
## Highlights

Journal Pre-proofs

- Phylogenomic tree of brown algae inferred from 32 nuclear protein-coding sequences in 39 Phaeophyceaean species.
- The new phylogeny is nearly fully supported and significantly improves on previous studies.
- Present results highlight the merit of nuclear-based genome data in solving interordinal phylogenetic relationships in brown algae.
- The new phylogeny provides a valuable tool for the study of the evolution of biological complexity in brown algae.

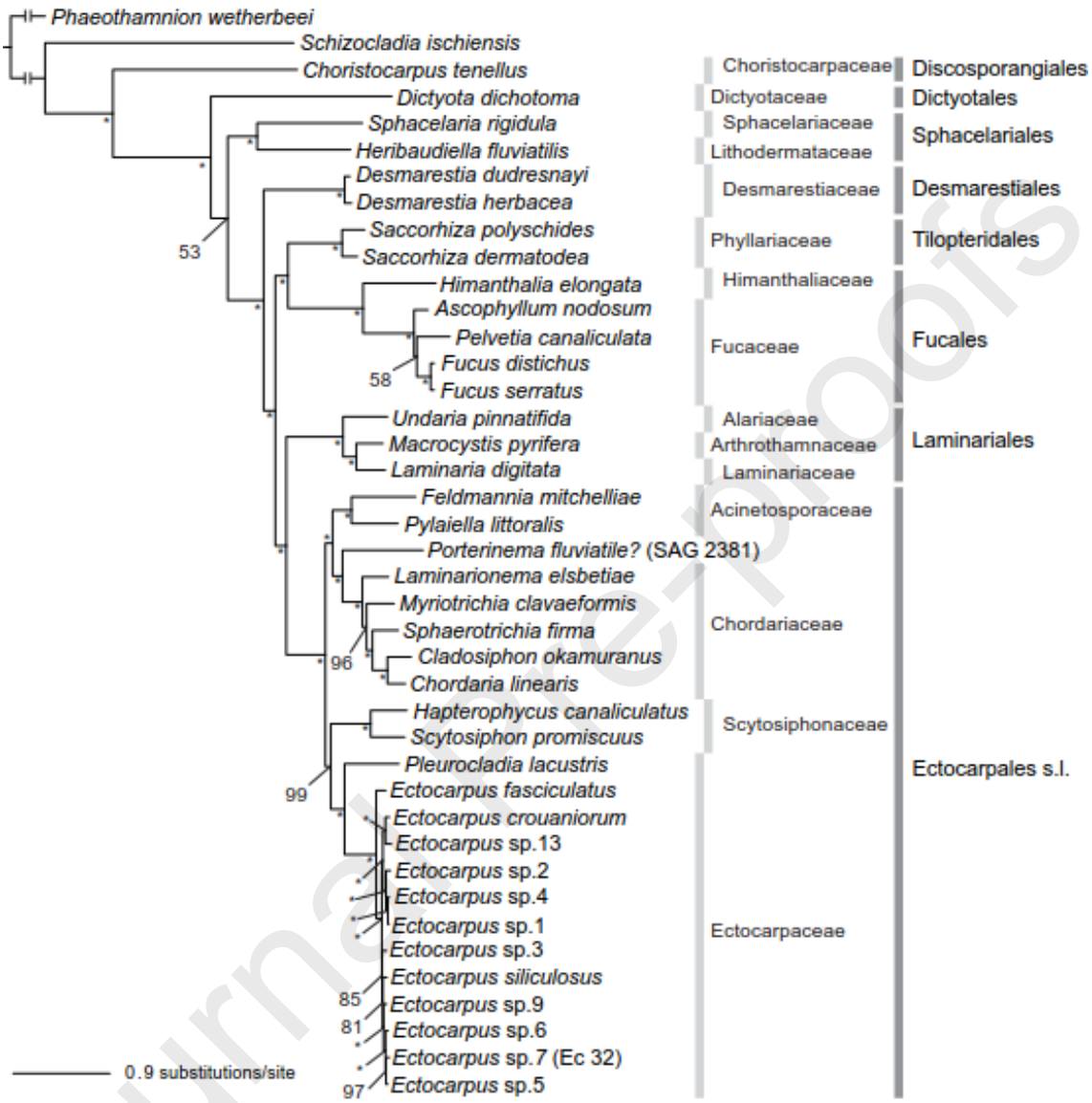
Journal Pre-proofs

Graphical Abstract




- ML tree constructed using RAxML-ng (53 OTUs, 32 genes, 15,597aa, 10,000 replicates of bootstrap analyses).
- Branches with a star indicate bootstrap value 100.

Journal Pre-proofs




- ML tree constructed using RAxML-ng (53 OTUs, 32 genes, 15,597aa, 10,000 replicates of bootstrap analyses).
- Branches with a star indicate bootstrap value 100.

Journal e-proofs