



The Rhodoexplorer Platform for Red Algal Genomics and Whole-Genome Assemblies for Several Gracilaria Species

Agnieszka P Lipinska, Stacy A Krueger-Hadfield, Olivier Godfroy, Simon M Dittami, Lígia Ayres-Ostrock, Guido Bonthond, Loraine Brillet-Guéguen, Susana Coelho, Erwan Corre, Guillaume Cossard, et al.

► To cite this version:

Agnieszka P Lipinska, Stacy A Krueger-Hadfield, Olivier Godfroy, Simon M Dittami, Lígia Ayres-Ostrock, et al.. The Rhodoexplorer Platform for Red Algal Genomics and Whole-Genome Assemblies for Several Gracilaria Species. *Genome Biology and Evolution*, 2023, 15 (7), pp.evad124. <10.1093/gbe/evad124>. <hal-04190579>

HAL Id: hal-04190579

<https://hal.sorbonne-universite.fr/hal-04190579v1>

Submitted on 29 Aug 2023

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.



HAL Authorization

The Rhodoexplorer Platform for Red Algal Genomics and Whole-Genome Assemblies for Several *Gracilaria* Species

Agnieszka P. Lipinska^{1,2,*†}, Stacy A. Krueger-Hadfield^{3,*†,‡}, Olivier Godfroy², Simon M. Dittami², Lígia Ayres-Ostrock^{4,5}, Guido Bonthond⁶, Loraine Brillet-Guéguen^{2,7}, Susana Coelho ¹, Erwan Corre⁷, Guillaume Cossard¹, Christophe Destombe⁸, Paul Epperlein¹, Sylvain Faugeton^{8,9}, Elizabeth Ficko-Blean², Jessica Beltrán^{8,9}, Emma Lavaut⁸, Arthur Le Bars^{7,10}, Fabiana Marchi⁴, Stéphane Mauger⁸, Gurvan Michel ², Philippe Potin², Delphine Scornet², Erik E. Sotka¹¹, Florian Weinberger¹², Mariana Cabral de Oliveira⁴, Marie-Laure Guillemin^{8,13,14}, Estela M. Plastino⁴, and Myriam Valero⁸

¹Department of Algal Development and Evolution, Max Planck Institute for Biology Tübingen, Tübingen, Germany

²CNRS, UMR 8227, Laboratory of Integrative Biology of Marine Models, Sorbonne Université, Station Biologique de Roscoff, Roscoff, France

³Department of Biology, University of Alabama at Birmingham, Alabama, USA

⁴Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, SP, Brazil

⁵Hortimare, Breeding and Propagating Seaweed, Heerhugowaard, The Netherlands

⁶Institute for Chemistry and Biology of the Marine Environment (ICBM), Carl von Ossietzky University Oldenburg, Wilhelmshaven, Germany

⁷CNRS, Sorbonne Université, FR2424, ABiMS-IFB, Station Biologique, Roscoff, France

⁸CNRS, Sorbonne Université, Pontificia Universidad Católica de Chile, Universidad Austral de Chile, IRL 3614, Evolutionary Biology and Ecology of Algae, Station Biologique de Roscoff, Roscoff, France

⁹Núcleo Milenio MASH, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile

¹⁰CNRS, Institut Français de Bioinformatique, IFB-core, Évry, France

¹¹Department of Biology, College of Charleston, Charleston, South Carolina, USA

¹²Marine Ecology Division, GEOMAR Helmholtz-Zentrum für Ozeanforschung, Kiel, Germany

¹³Núcleo Milenio MASH, Facultad de Ciencias, Instituto de Ciencias Ambientales y Evolutivas, Universidad Austral de Chile, Valdivia, Chile

¹⁴Centro FONDAP de Investigación de Ecosistemas Marinos de Altas Latitudes (IDEAL), Valdivia, Chile

[†]These authors shared first authorship.

[‡]Present address: Virginia Institute of Marine Science, Eastern Shore Laboratory, 40 Atlantic Ave, Wachapreague, VA 23480, USA.

*Corresponding authors: E-mails: alipinska@tuebingen.mpg.de; sakh@vims.edu.

Accepted: 29 June 2023

Abstract

Macroalgal (seaweed) genomic resources are generally lacking as compared with other eukaryotic taxa, and this is particularly true in the red algae (Rhodophyta). Understanding red algal genomes is critical to understanding eukaryotic evolution given that red algal genes are spread across eukaryotic lineages from secondary endosymbiosis and red algae diverged early in the Archaeplastids. The Gracilariales is a highly diverse and widely distributed order including species that can serve as ecosystem engineers in intertidal habitats and several notorious introduced species. The genus *Gracilaria* is cultivated worldwide, in part for its production of agar and other bioactive compounds with downstream pharmaceutical and industrial applications. This genus is also emerging as a model for algal evolutionary ecology. Here, we report new whole-genome assemblies for two species (*Gracilaria chilensis* and *Gracilaria gracilis*), a draft genome assembly of *Gracilaria caudata*, and genome annotation

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

of the previously published *Gracilaria vermiculophylla* genome. To facilitate accessibility and comparative analysis, we integrated these data in a newly created web-based portal dedicated to red algal genomics (<https://rhodoexplorer.sb-roscoff.fr>). These genomes will provide a resource for understanding algal biology and, more broadly, eukaryotic evolution.

Key words: evolution, ecology, omics, ploidy, Rhodophyta.

Significance

The Gracilariales are an ecologically and economically important red algal order found throughout the coastal regions of the world. Understanding the biology, ecology, and evolution of species in this order, and that of red algae more broadly, has been hampered by the limited phylogenetic coverage of genomic resources. Here, we present whole-genome assemblies and gene annotations for four *Gracilaria* species that will serve as a key resource for algal research on evolution, ecology, biotechnology, and aquaculture.

Introduction

Red algae (Rhodophyta) represent a lineage of photosynthetic eukaryotes in the Archaeplastids that diverged from green algae around 1,700 Ma (Yang et al. 2016). Within the Rhodophyta, the Cyanidiophyceae were the earliest to diverge ~1,200 Ma, while the Florideophyceae diverged more recently (i.e., 412 Ma; Yang et al. 2016) and constitute the most speciose group (Graham et al. 2016). In this context, the genomic resources currently available (supplementary table S1, Supplementary Material online) represent only a fraction of the diversity of red algae, limiting our capacity to reconstruct the evolutionary history of the unique features of this group.

The Florideophyceae have a life cycle in which haploid male and female gametophytes alternate with a diploid tetrasporophyte (but see supplementary fig. S1, Supplementary Material online). Many species have “isomorphic” gametophytes and tetrasporophytes, which are hard to discern without the aid of molecular tools (e.g., sex-linked markers, Martinez et al. 1999; Guillemain et al. 2012; or microsatellites, Krueger-Hadfield et al. 2016).

Here, we focus on four *Gracilaria* (There is controversy over the systematics of *Gracilaria* Greville, but for the purposes of this paper, we consider the four species as belonging to the genus *Gracilaria* [sensu Lyra et al. 2021; Guiry and Guiry 2022]). species spanning roughly 170 Myr of evolution (Lyra et al. 2021). These species were chosen based on their evolutionary, ecological, and/or economic importance. Species in the genus *Gracilaria* produce agars in their cell wall (Popper et al. 2011); they can be propagated vegetatively and serve as ecosystem engineers in intertidal zone (Kain and Destombe 1995). The four taxa chosen can be divided into three clades based on their molecular divergence: 1) *Gracilaria chilensis* and *Gracilaria vermiculophylla*, 2) *Gracilaria caudata*, and 3) *Gracilaria gracilis* (Lyra et al. 2021). *Gracilaria gracilis* and *G. caudata* are evolutionarily more distinct than the phylogenetic group that contains

G. chilensis and *G. vermiculophylla*. *Gracilaria chilensis* C.J. Bird et al. is an important crop along the Chilean coastline, where it has been both harvested and subsequently planted after a crash in natural stands likely due to overharvesting (Buschmann et al. 2001). The artificial selection for tetrasporophytes has resulted in early stages of domestication (Valero et al. 2017) and loss of sexual reproduction (Guillemain et al. 2008). *Gracilaria vermiculophylla* (Ohmi) Papenfuss is a successful invader in many of the bays and estuaries of the Northern Hemisphere (Krueger-Hadfield et al. 2017). These invasions were likely facilitated by adaptive shifts in temperature and salinity tolerance (e.g., Sotka et al. 2018) and to biofoulers (e.g., Bonthond et al. 2020), as well as the ability to fragment (Krueger-Hadfield et al. 2016). *Gracilaria caudata* J. Agardh can form dense stands in the intertidal zone (Plastino and Oliveira 1997) and has been subjected to intense harvesting pressure, leading to declines in native populations (Hayashi et al. 2014; see also Ayres-Ostrock et al. 2019). Finally, *G. gracilis* (Stackhouse) Steentoft, L.M. Irvine & Farnham is a long-lived species that inhabits tide pools along European coastlines. This species serves as model species to test hypotheses related to the evolution of sex (e.g., alternation of haploid and diploid phases in life cycles, Destombe et al. 1989, 1992, 1993; Hughes and Otto 1999; mating system and sexual selection, Richerd et al. 1993; Engel et al. 1999).

The availability of genomic and genetic resources for these four *Gracilaria* species should aid in our understanding of the evolutionary ecology of red algae in their dynamic environment, during invasions of new habitats, under cultivation practices, and in response to climate change. Moreover, these new resources will add to the existing genomic data and illuminate key processes in eukaryotic evolution. The Rhodoexplorer Red Algal Genome Database currently includes the *Gracilaria* species discussed here but will include all the high-quality genomic resources available for the Rhodophyta (e.g., genomes and transcriptomes), thereby providing a unique resource for comparative analyses.

Table 1.

Assembly Statistics

	<i>G. chilensis</i>	<i>G. vermiculophylla</i>	<i>G. caudata</i>	<i>G. gracilis</i>
Strain	NLEC103-M9	HapMaleFtJ-2017	M-176_S67	GNS1m
Sequencing	PacBio	Illumina, HiC	Illumina	PacBio
Genome size	76.07 Mb	44.95 Mb	30.28 Mb	72.49 Mb
Contigs/scaffolds	138/138	7,753/4,240	55,767/5,535	279/279
GC contents	48.9%	49.5%	49.9%	46.6%
N50	1.56 Mb	2.56 Mb	20.8 kb	563 kb
L50	18	6	396	38
Repeat content	66.2%	48.3%	45.7%	60.7%
Protein-coding genes	7,943	6,807	8,737	9,460
Av. gene length	1,404 bp	1,751 bp	1,409 bp	1,643 bp
Genes w. interpro/Uniprot 90 ^a	93.4%/88.8%	93.6%/89.7%	91.7%/86.5%	92.0%/84.2%
Genes with GO annotation	52.7%	54.4%	49.9%	47.9%
Genes with intron	23.4%	24.1%	28.6%	29.4%
BUSCO complete	75.3%	65.1%	73.0%	77.3%
BUSCO fragmented	6.3%	6.7%	8.6%	6.3%
BUSCO missing	18.4%	28.2%	18.4%	16.4%

^ae-value cutoff 1e-5.

Results and Discussion

Genome Assembly

Genome assembly sizes were 72 and 76 Mb for *G. gracilis* and *G. chilensis*, respectively. In addition, we created a draft genome assembly based on the Illumina sequencing only for *G. caudata* (30 Mb) and reassembled the genome of *G. vermiculophylla* (Flanagan et al. 2021) to a final 45 Mb after bacterial contamination removal. The above genome assemblies were comparable with the genomes of *Gracilaria domingensis* (78 Mb; Nakamura-Gouvea et al. 2022) and *Gracilaria changii* (36 Mb; Ho et al. 2018). PacBio assemblies of *G. chilensis* and *G. gracilis* produced here (138 and 279 contigs, respectively; N50 of 1.56 and 0.56 Mb, respectively) are the most contiguous red macroalgal genomes presently available in public databases, apart from *G. vermiculophylla* and *Pyropia yezoensis* where the addition of a HiC library enabled scaffolding nearly at the chromosome level (Wang et al. 2020; Flanagan et al. 2021). In *G. vermiculophylla*, however, regardless of the high N50 of 2.56 Mb, the total number of contigs/scaffolds was also high (7,753/4,240). The *G. caudata* assembly was fragmented with a low N50 of 21 kb and 55,767/5,535 contigs/scaffolds. Despite the differences in assembly size, BUSCO scores were similar across the long-read-sequenced *G. gracilis* and *G. chilensis* (83.6% and 81.6% of conserved proteins present) and the more fragmented *G. caudata* genome (81.6%, Eukaryota_odb10; Manni et al. 2021, Simão et al. 2015; table 1). The reassembled genome of *G. vermiculophylla* contained 71.8% of the conserved proteins. Given the diversity of Rhodophyta and the lack of lineage-specific databases, these results are in the expected range. A recent study estimated the presence of

conserved eukaryotic genes (Eukaryota_odb10) in red algal genomes at a median level of 69% (Hanschen et al. 2020).

Red algal genomes are repeat rich, with half or more of their genomic sequence being constituted by repetitive elements, as reported previously for *Porphyra umbilicalis* (43.9%; Brawley et al. 2017), *P. yezoensis* (48%; Wang et al. 2020), and *Chondrus crispus* (73%; Collén et al. 2013). In agreement with this general trend, between 45.7% and 66.2% of the *Gracilaria* genomes corresponded to repetitive elements (fig. 1, supplementary fig. S2, Supplementary Material online, and table 1).

Gene Prediction and Annotation

Gene prediction yielded a total of 7,943, 8,737, and 9,460 protein-coding sequences for *G. chilensis*, *G. caudata*, and *G. gracilis* (table 1), which was comparable with other red macroalgal genomes, *C. crispus* (9,815 genes; Collén et al. 2013) and *G. changii* genome (10,912 genes; Ho et al. 2018). In addition, we annotated the reassembled genome of *G. vermiculophylla*, which yielded fewer genes (6,807). Among these genes, 70.6–76.6% did not contain any introns, as typical for the compact genomes of red algae (Qiu et al. 2015). Most *Gracilaria* genes had homologous sequences in the Uniprot database (84.2–89.7%) and were annotated with at least one INTERPRO hit (91.7–93.6%). Between 47.9% and 54.4% of genes were associated with gene ontology (GO) annotations.

OrthoFinder analyses identified 4,666 orthogroups present in all four genomes (supplementary fig. S2, Supplementary Material online) versus 408–620 orthogroups or orphan genes specific to only one of the sequenced species (supplementary fig. S2, Supplementary Material online). Among the species-specific sequences, the rate of GO

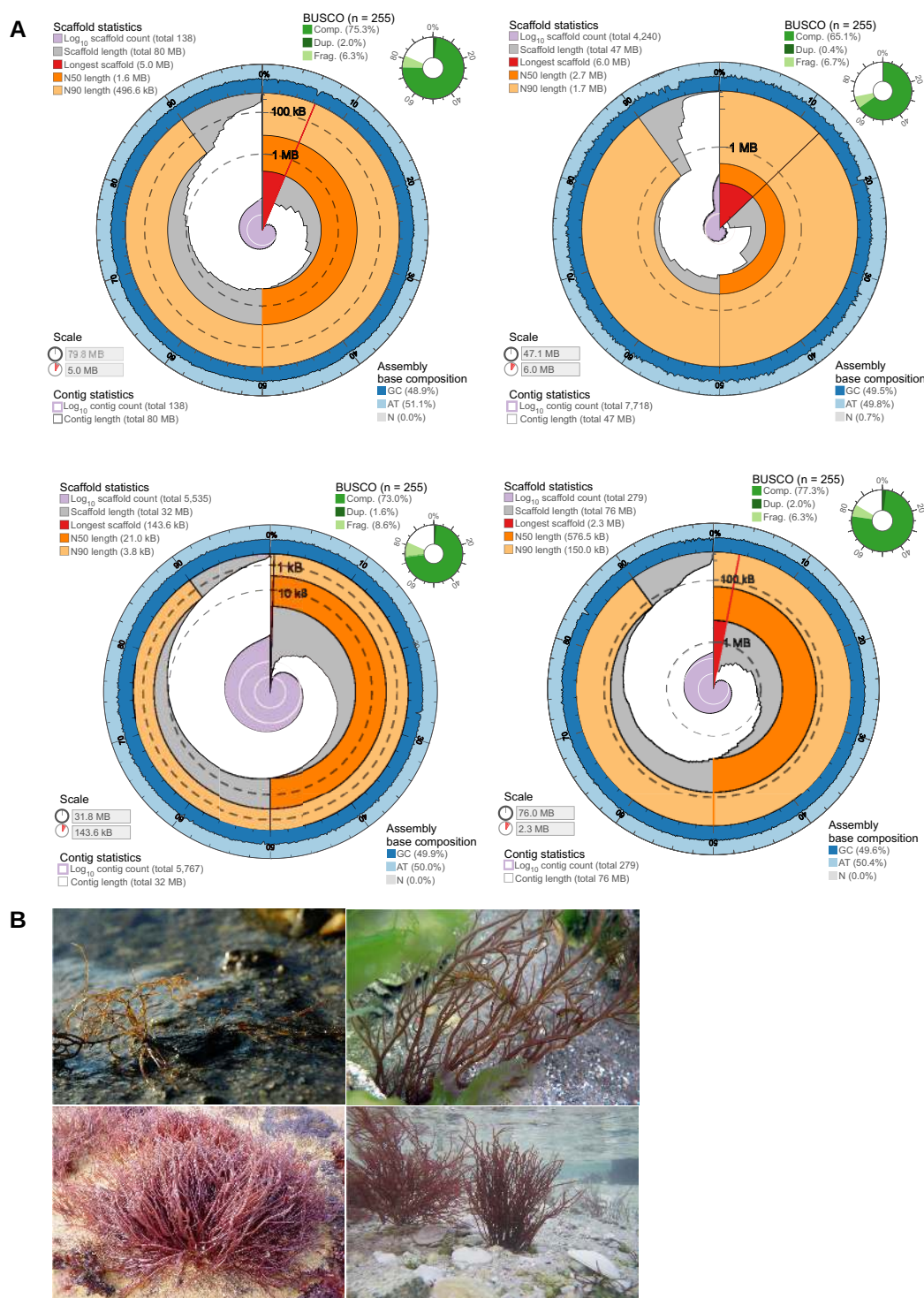


FIG. 1.—(A) Genome assembly metrics of *G. chilensis* (top left), *G. vermiculophylla* (top right), *G. caudata* (bottom left), and *G. gracilis* (bottom right) (Challis 2017; <https://github.com/rjchallis/assembly-stats>). The inner radius of the circular plot represents the length of the longest scaffold in the assembly and the proportion of the assembly that it represents. The cumulative number of scaffolds within a given percentage of the genome is plotted in light purple originating at the center of the plot. The N50 and N90 scaffold lengths are indicated by dark and light orange, respectively. Genome scaffolds are plotted in gray from the circumference and the length of segment at a given percentage indicates the cumulative percentage of the assembly that is contained within scaffolds of at least that length. The guanine–cytosine (GC) content is marked by the dark blue outer circle. Complete, fragmented, and duplicated BUSCO genes are shown in green in the upper right corner. (B) *Gracilaria chilensis* (top left), *G. vermiculophylla* (top right), *G. caudata* (bottom left), and *G. gracilis* (bottom right). Photo credit in order: M.-L. Guillemin, S. Krueger-Hadfield, E. M. Plastino, C. Destombe.

annotation was lower than for the entire data set, ranging from 12.7% for *G. chilensis* to 18.2% for *G. caudata*. Both the annotated and the unknown species-specific genes constitute attractive targets to study their role in adaptation and speciation.

Rhodoexplorer Red Algal Genome Database

In addition to depositing the raw reads and sequenced genome in a public repository, we integrated the data into the newly created Rhodoexplorer Red Algal Genome Database (<https://rhodoexplorer.sb-roscoff.fr>), which will include more red algal genomes in the future. The services provided include the following:

- Information about the sequenced strains, with links to external databases (NCBI, WoRMS, and Algaebase).
- Assembly and annotation metrics.
- Data downloads: genomic, genes and proteomic data sets, structural and functional annotations, orthology clusters, etc.
- A Blast interface with a selection of red algal genomes, predicted and de novo assembled transcriptomes and proteomes.
- Visualization tools: a genome browser to visualize the predicted genes and the RNA-sequence (RNAseq) data mapped on the genome and a web interface to visualize functional annotations and retrieve individual protein sequences.

Materials and Methods

Sampling of the Biological Material

Adult female and male *Gracilaria* thalli, all bearing reproductive structures, were collected from natural populations: *G. chilensis* in Lenca (Chile, −41.607, −72.692), *G. vermiculophylla* in Charleston, SC (USA, 32.752, −79.900), *G. caudata* in Paracuru, CE (Brazil, −3.399, −39.012), and *G. gracilis* in Cape Gris-Nez (France, 50.872, 1.584). *Gracilaria caudata* and *G. chilensis* were maintained as clonal, unialgal cultures under laboratory conditions prior to nucleic acid extractions (see *Culture conditions*). Field-collected *G. gracilis* and *G. vermiculophylla* thalli were transported to the laboratory, examined under a microscope, and cleaned of contaminants. If visible, cystocarps were excised prior to preservation of the thalli at −80°C. [Supplementary table S2, Supplementary Material online](#) provides details of the *Gracilaria* species used in this study.

Culture Conditions

Cultures were initiated either from lab crosses or from tetraspores released by field-collected tetrasporophytes. *Gracilaria caudata* was grown in the modified von Stosch

nutrient solution (Ursi and Plastino 2001) diluted to 25% in seawater (32 practical salinity unit [psu]), with weekly renewals. The algae were kept in culture chambers at 25 °C under fluorescent illumination of 70 $\mu\text{mol m}^{-2} \text{s}^{-1}$ 14-h photoperiod, following previously established optimal growth conditions (Yokoya and Oliveira 1992a, 1992b). *Gracilaria chilensis* was grown in Provasoli medium (McLachlan 1973), changed weekly during the first 2 months and twice a week thereafter. Cultures were kept at 13 °C under 40–60 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of light with 12-h day length.

Nucleic Acid Extraction, Library Preparation, and Sequencing

Genomic DNA (gDNA) was extracted using DNeasy PowerPlant Pro Kit for *G. caudata* or an in-house protocol based on Faugeron et al. (2001) for *G. chilensis* and *G. gracilis*. The concentration and purity of DNA were measured with NanoDrop and Qubit before sequencing on an Illumina HiSeq 2500 (125-bp PE reads for *G. chilensis* and *G. gracilis*; 100-bp PE reads for *G. caudata*) or PacBio Sequel II with sheared gDNA large insert library (*G. gracilis* and *G. chilensis*) ([supplementary table S2, Supplementary Material online](#)).

For genome annotation, total RNA was extracted from mature thalli of male and female gametophytes of *G. chilensis*, *G. caudata*, and *G. gracilis* using the RNeasy Mini Plant Kit (Qiagen) and Macherey Nagel NucleoSpin RNA Plant Kit for *G. vermiculophylla*, following the manufacturer's instructions. Paired-end 150-bp Illumina reads were generated with Illumina HiSeq 2500 ([supplementary table S2, Supplementary Material online](#)).

Genome Assembly

De novo genome assemblies for *G. gracilis* and *G. chilensis* were generated based on 203-fold and 116-fold coverage of PacBio long reads, respectively. Bacterial sequences were removed from raw data (subreads) using BlobTools v1.1.1 (Laetsch and Blaxter 2017). Two independent assemblies were generated using CANU (Koren et al. 2017) and FLYE (Kolmogorov et al. 2019). Based on congruity (QUAST v.5.0.2; Mikheenko, et al. 2018) and BUSCO score (Simão et al. 2015), the best assembly was kept and polished using three iterations of RACON v.1.4.20. Finally, PacBio sequencing error was corrected using 150-bp paired-end Illumina reads with PILON v.1.23 software (Walker et al. 2014). The draft genome assembly of *G. caudata* was generated using 171-fold coverage of 150-bp paired-end Illumina reads only. First, a meta-genome was produced using metaSPAdes v3.12.0 (Nurk et al. 2017) and bacterial contigs were detected using BlobTools. Reads corresponding to eukaryotic contigs were then assembled using SPAdes v3.12.0 (Bankevich et al. 2012).

For *G. vermiculophylla*, we updated the existing chromosome-scale genome assembly (Flanagan et al. 2021) by reassembling the Illumina reads using SPAdes v3.12.0 (Bankevich et al. 2012) and scaffolding with HiC libraries, following the Dovetail Genomics proprietary pipeline (Elbers et al. 2019). This process ameliorated the genome continuity (N50 increased from 2.06 to 2.68 Mb) and completeness (BUSCO score increased from 57.6% to 65.9% of complete genes using the Eukaryota_odb10 data set).

Genome assemblies were validated with a final BlobTools v1.1.1 analysis (Laetsch and Blaxter 2017) using DNAseq mapping coverage files produced by HISAT2 v2.2.1 (Kim et al. 2019), Diamond BlastX v2.0.11 (Buchfink et al. 2015, 2021) hit-file against nonredundant protein sequences archive from NCBI (-sensitive, -max-target-seqs 1, -e-value 1e-20), and Blast v2.12.0 (Camacho et al. 2009) output against nucleotide archive from NCBI (-max_target_seqs 10 -max_hsps 1 -evalue 1e-20) as input genomic scaffolds classified as bacterial or with a coverage of <1 (sum of coverages for each sequence across all coverage files) were removed from the assembly. Genome assembly completeness was assessed using BUSCO scores with the eukaryotic data set (Eukaryota_odb10; Simão et al. 2015; Manni et al. 2021).

Chloroplastic and mitochondrial genomes of each species were reconstructed from Illumina raw reads using NOVOPlasty (Dierckxsens et al. 2017) through the European Galaxy web portal (<https://usegalaxy.eu/>). Annotation of those de novo organellar genomes was done using the GeSeq web tool (Tillich et al. 2017; <https://chlorobox.mpimp-golm.mpg.de/geseq.html>). Public sequences from *G. caudata* voucher SPF:57390 (NC_039146, NC_039139), *G. chilensis* voucher CNU050183 (KP728466, KT266788), *G. gracilis* voucher SPF:55734 (NC_039141, NC_039148), and *G. vermiculophylla* (MN853882, MH396022) were retrieved from NCBI and used as seeds and references for both assembly and annotation.

Genome Annotation

Each reference genome was first masked using RepeatMasker v4.0.9 (Smit et al. 2013–2015) with Dfam v3.0 database (Wheeler et al. 2013) and a customized repeat library produced from concatenated outputs of RepeatScout v1.0.6 (Price et al. 2005) and TransposonPSI v1.0.0 (Haas 2007–2011). Initial quality assessment of the RNAseq reads was performed with FastQC v0.11.9 (Andrews 2010), and reads were trimmed using Trimmomatic v0.39 (TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50; Bolger et al. 2014). Clean reads were mapped to the reference genome assembly using HISAT2 v2.2.1 (Kim et al. 2019) and used to annotate protein-coding genes with BRAKER2 v2.1.6 (Bruna et al. 2021). Functional annotation of the transcriptomes was performed using

eggNOG-mapper (Huerta-Cepas et al. 2019; Cantalapiedra et al. 2021).

All codes used for genomes assembly and annotation are available on the Gitpage dedicated to the genome database project https://abims-sbr.gitlab.io/rhodoexplorer/doc/data_process/.

Rhodoexplorer Red Algal Genome Database

The main web portal (<https://rhodoexplorer.sb-roscoff.fr>) has been implemented using the Python web framework Django, with data stored in a relational database (PostgreSQL).

For each red algal species, an integrated environment of visualization tools has been deployed based on the Galaxy Genome Annotation (GGA) project (Bretaud et al. 2019). Each GGA environment deployed for the Rhodoexplorer database includes the following: Chado, a PostgreSQL relational database schema for storing biological data (Mungall et al. 2007); JBrowse, a web-based genome browser (Buels et al. 2016); Tripal, a Drupal-based application for creating biological websites (Sanderson et al. 2013); Elasticsearch, a distributed, free, and open search and analytics engine for all types of data (<https://www.elastic.co/products/elasticsearch>); and Galaxy, a browser-accessible workbench for scientific computing used as a data loading orchestrator for administrators (The Galaxy Community 2022). To facilitate the deployment and the administration of the GGA service, a set of Python tools has been developed (http://gitlab.sb-roscoff.fr/abims/e-infra/gga_load_data) allowing mass deployment of Docker containers and automated data loading through Galaxy with the Bioblend API (Sloggett et al. 2013).

The Blast interface (<https://blast.sb-roscoff.fr/rhodoexplorer/>) includes an implementation of the Blast algorithm using SequenceServer (Priyam et al. 2019) graphical.

The documentation website for navigating the platform web portal and resources (<https://abims-sbr.gitlab.io/rhodoexplorer/doc/>) is published from a GitLab repository, with Pages and MkDocs, a static site generator.

The entire informatic infrastructure is deployed and maintained on the ABiMS Bioinformatics platform of the Roscoff Biological Station, part of the national infrastructure French Bioinformatic Institute.

Supplementary material

Supplementary data are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

This project was supported by start-up funds from the College of Arts and Sciences at the University of Alabama

at Birmingham to S.A.K.-H.; ANID NCN2021-033 and FONDECYT 1221456 and 1221477 to M.L.G., J.B., and S.F.; the International Research Networks DEBMA “Diversity, Evolution and Biotechnology of Marine Algae” (CNRS GDRI 0803) and DABMA “Diversity, Adaptation, and Biotechnology of Marine Algae” (CNRS IRN 00022); the ERC (grant number 864038 to S.M.C.); and the ANR project IDEALG (ANR-10-BTBR-04, “Investissements d’Avenir, Biotechnologies-Bioressources”). We are grateful to the Roscoff Bioinformatics platform ABiMS (<http://abims.sb-roscoff.fr>), part of the Institut Français de Bioinformatique (ANR-11-INBS-0013) and BioGenouest network, and the Max Planck Institute for Biology Tübingen for providing computational resources. We also wish to thank Kristy Hill-Spanik, Rosário Petti, and Vivian Viana for field and technical support.

Data Availability

Sequencing data have been deposited in the SRA database under BioProjects PRJNA936482, PRJNA931233, PRJNA938301, and PRJNA938403. The accession numbers for the raw sequence data are provided in [supplementary table S2, Supplementary Material](#) online.

Gracilaria chilensis, *G. gracilis*, and *G. caudata* Whole Genome Shotgun project have been deposited at DDBJ/ENA/GenBank under the accessions JARGXX0000000000, JARGSG0000000000, and JASCIV0000000000, respectively. *Gracilaria vermiculophylla* updated assembly has been deposited under JAHNZQ0000000000.

Literature Cited

- Andrews. FastQC: a quality control tool for high throughput sequence data [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. 2010
- Ayres-Ostrock LM, et al. 2019. Dual influence of terrestrial and marine historical processes on the phylogeography of the Brazilian intertidal red alga *Gracilaria caudata*. *J Phycol.* 55:1096–1114.
- Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comp Biol.* 19: 455–477.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Bonthond G, et al. 2020. How do microbiota associated with an invasive seaweed vary across scales? *Mol Ecol.* 29:2094–2108.
- Brawley SH, et al. 2017. Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangiophyceae, Rhodophyta). *PNAS* 114:E6361–E6370.
- Bretaudeau A, et al. 2019. Galaxy Genome Annotation: easier genome annotation using Galaxy and GMOD tools [version 1; not peer reviewed]. *F1000Res.* 8:1026.
- Bruna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform.* 3(1):lqaa108.
- Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Meth.* 18:366–368.
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Meth.* 12:59–60.
- Buels R, et al. 2016. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* 17:66.
- Buschmann AH, Correa JA, Westermeier R, Hernández-González MDC, Norambuena R. 2001. Red algal farming in Chile: a review. *Aquaculture* 194:203–220.
- Camacho C, et al. 2009. Blast+: architecture and applications. *BMC Bioinform.* 10:421.
- Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol.* 38:5825–5829.
- Challis R. 2017 Assembly-stats, Assembly statistic visualisation. <https://github.com/rjchallis/assembly-stats/tree/17.02>
- Collén J, et al. 2013. Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proc Natl Acad Sci U S A.* 110:5247–5252.
- Destombe C, Godin J, Lefebvre C, Dehorter O, Vernet P. 1992. Differences in dispersal abilities of haploid and diploid spores of *Gracilaria verrucosa* (Gracilariales, Rhodophyta). *Bot Mar.* 35:93–98.
- Destombe C, Godin J, Nocher M, Richerd S, Valero M. 1993. Differences in response between haploid and diploid isomorphic phases of *Gracilaria verrucosa* (Rhodophyta: Gigartinales) exposed to artificial environmental conditions. *Hydrobiologia* 260:131–137.
- Destombe C, Valero M, Vernet P, Couvet D. 1989. What controls haploid-diploid ratio in the red alga, *Gracilaria verrucosa*? *J Evol Biol.* 2:317–338.
- Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45:e18.
- Elbers JP, et al. 2019. Improving Illumina assemblies with Hi-C and long reads: an example with North African dromedary. *Mol Ecol Res.* 19: 1015–1026.
- Engel CR, et al. 1999. Performance of non-motile male gametes in the sea: analysis of paternity and fertilization success in a natural population of a red seaweed, *Gracilaria gracilis*. *Proc R Soc Lond B Biol Sci.* 266:1879–1886.
- Faugeron S, Valero M, Destombe C, Martinez EA, Correa JA. 2001. Hierarchical spatial structure and discriminant analysis of genetic diversity in the red alga *Mazzaella laminarioides* (Gigartinales, Rhodophyta). *J Phycol.* 37:705–716.
- Flanagan BA, et al. 2021. Founder effects shape linkage disequilibrium and genomic diversity of a partially clonal invader. *Mol Ecol.* 30: 1962–1978.
- The Galaxy Community. 2022. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* 50:W345–W351.
- Graham LE, et al. 2016. Algae. Madison: LJLM Press. p. 720.
- Guillemin ML, et al. 2008. Genetic variation in wild and cultivated populations of the haploid-diploid red alga *Gracilaria chilensis*: how farming practices favor asexual reproduction and heterozygosity. *Evol.* 62:1500–1519.
- Guillemin ML, Huanel OR, Martinez EA. 2012. Characterization of genetic markers linked to sex determination in the haploid-diploid red alga *Gracilaria chilensis*. *J Phycol.* 48:365–372.
- Guiry MD, Guiry GM 2022. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. <https://www.algaebase.org>
- Haas BJ 2007–2011. TransposonPSI. <http://transposonpsi.sourceforge.net>
- Hanschen ER, Hovde BT, Starkenburg SR. 2020. An evaluation of methodology to determine algal genome completeness. *Algal Res.* 51:102019.
- Hayashi L, Bulboa C, Kradolfer P, Soriano G, Robledo D. 2014. Cultivation of red seaweeds: a Latin American perspective. *J Appl Phycol.* 26:719–727.

- Ho C-L, Lee W-K, Lim E-L. 2018. Unraveling the nuclear and chloroplast genomes of an agar producing red macroalga, *Gracilaria changii* (Rhodophyta, Gracilariiales). *Genomics* 110:124–133.
- Huerta-Cepas J, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acid Res.* 47:D309–D314.
- Hughes JS, Otto SP. 1999. Ecology and the evolution of biphasic life cycles. *Am Nat.* 154:306–320.
- Kain JM, Destombe C. 1995. A review of the life history, reproduction and phenology of *Gracilaria*. *J Appl Phycol.* 7:269–281.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 37:907–915.
- Kolmogorov M, Yuan J, Lin Y, Yevzner P. 2019. Assembly of long error-prone reads using repeat graphs. *Nat Biotechnol.* 37:540–546.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, et al. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27:722–736.
- Krueger-Hadfield SA, et al. 2016. Invasion of novel habitats uncouples haplo-diplontic life cycles. *Mol Ecol.* 25:3801–3816.
- Krueger-Hadfield SA, et al. 2017. Genetic identification of source and likely vector of a widespread marine invader. *Ecol Evol.* 7:4432–4447.
- Laetsch DR, Blaxter ML. 2017. KinFin: software for taxon-aware analysis of clustered protein sequences. *G3 (Bethesda).* 7:3349–3357.
- Lyra GDM, et al. 2021. Phylogenomics, divergence time estimation and trait evolution provide a new look into the Gracilariiales (Rhodophyta). *Mol Phylogenetics Evol.* 165:107294.
- Manni M, et al. 2021. BUSCO: assessing genomic data quality and beyond. *Current Protocols.* <https://currentprotocols.onlinelibrary.wiley.com/doi/full/10.1002/cpz1.323>
- Martinez EA, Destombe C, Quillet MC, Valero M. 1999. Identification of random amplified polymorphic DNA (RAPD) markers highly linked to sex determination in the red alga *Gracilaria gracilis*. *Mol Ecol.* 8:1533–1538.
- McLachlan J. 1973. Growth media—marine. In: Stein JR, editor. *Handbook of phycological methods, culture methods and growth measurements*. Cambridge: Cambridge University Press. p. 25–52.
- Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34(13):i142–i150.
- Mungall CJ, Emmert DB; FlyBase Consortium. 2007. A Chado case study: an ontology-based modular schema for representing genome-associated biological information. *Bioinformatics* 23:i337–i346.
- Nakamura-Gouveia N, et al. 2022. Insights into agar and secondary metabolite pathways from the genome of the red alga *Gracilaria domingensis* (Rhodophyta, Gracilariiales). *J Phycol.* 58:406–423.
- Nurk S, Meleshko D, Korobeynikov A, Pevzner P. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27:824–834.
- Plastino EM, Oliveira EC. 1997. *Gracilaria caudata* J. Agardh (Gracilariiales, Rhodophyta)—restoring an old name for a common western Atlantic alga. *Phycologia* 36:225–332.
- Popper ZA, et al. 2011. Evolution and diversity of plant cell walls: from algae to flowering plants. *Ann Rev Plant Biol.* 62:567–590.
- Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. To appear in *Proceedings of the 13 Annual International conference on Intelligent Systems for Molecular Biology (ISMB-05)*. Detroit, Michigan.
- Priyam A, et al. 2019. Sequenceserver: a modern graphical user interface for custom BLAST databases. *Mol Biol Evol.* 36:2922–2924.
- Qiu H, Price DC, Yang EC, Yoon HS, Bhattacharya D. 2015. Evidence of ancient genome reduction in red algae (Rhodophyta). *J Phycol.* 51: 624–636.
- Richerd S, Destombe C, Cuguen J, Valero M. 1993. Variation of reproductive success in a haplo-diploid red alga, *Gracilaria verrucosa*: effects of parental identities and crossing distance. *Am J Bot.* 80:1379–1391.
- Sanderson LA, et al. 2013. Tripal v1.1: a standards-based toolkit for construction of online genetic and genomic databases. *Database* 2013: bat075.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212.
- Sloggett C, Goonasekera N, Afgan E. 2013. Bioblend: automating pipeline analyses within Galaxy and CloudMan. *Bioinformatics* 29:1685–1686.
- Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0. 2013–2015 <http://www.repeatmasker.org>
- Sotka EE, et al. 2018. Combining niche shift and population genetic analyses predicts rapid phenotypic evolution during invasion. *Evol Appl.* 11:781–793.
- Tillich M, et al. 2017. Geseq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45:W6–W11.
- Ursi S, Plastino EM. 2001. Crescimento in vitro de linhagens de coloração vermelha e verde clara de *Gracilaria birdiae* (Gracilariiales, Rhodophyta) em dois meios de cultura: análise de diferentes estádios reprodutivos. *Rev. Brasil. Bot.* 24:585–592.
- Valero M, et al. 2017. Perspectives on domestication research for sustainable seaweed aquaculture. *Perspectives in Phycology.* 4: 33–46.
- Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11):e112963.
- Wang D, et al. 2020. *Pyropia yezoensis* genome reveals diverse mechanisms of carbon acquisition in the intertidal environment. *Nat Commun.* 11:4028.
- Wheeler TJ, et al. 2013. Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Res.* 41:D70–D82.
- Yang EC, et al. 2016. Divergence time estimates and the evolution of major lineages in the florideophyte red algae. *Sci Rep.* 6:21361.
- Yokoya NS, Oliveira EC. 1992a. Effects of salinity on the growth rate, morphology and water content of some Brazilian red algae of economic importance. *Cienc Mar.* 18:49–64.
- Yokoya NS, Oliveira EC. 1992b. Temperature response of economically important red algae and their potential for mariculture in Brazilian waters. *J Appl Phycol.* 4:339–345.

Associate editor: Vincent Castric