J. Phycol. **51,** 586–589 (2015)

DOI: 10.1111/jpy.12297

NOTE

THE PLASTID GENOME OF THE RED ALGA *LAURENCIA*[1]

Heroen Verbruggen[2] *and Joana F. Costa*

School of BioSciences, University of Melbourne, Melbourne, Victoria 3010, Australia

We present the 174,935 nt long plastid genome of the red alga *Laurencia* sp. JFC0032. It is the third plastid genome characterized for the largest order of red algae (Ceramiales). The circular-mapping plastid genome is small compared to most florideophyte red algae, and our comparisons show a trend toward smaller plastid genome sizes in the family Rhodomelaceae, independent from a similar trend in Cyanidiophyceae. The *Laurencia* genome is densely packed with 200 annotated protein-coding genes (188 widely conserved, 3 open reading frames shared with other red algae and 9 hypothetical coding regions). It has 29 tRNAs, a single-copy ribosomal RNA cistron, a tmRNA, and the RNase P RNA.

Key index words: **Ceramiales; *Laurencia*; plastid genome; Rhodomelaceae; Rhodophyta**

***Abbreviations*: CDS, coding sequence; CTAB, cetyltrimethylammonium bromide; ORF, open reading frame**

The fact that red algae are underrepresented in data sets of plastid genomes has been highlighted in recent work as a major gap in our knowledge of plastid evolution (Lang and Nedelcu 2012, Janouškovec et al. 2013). This situation is improving thanks to high-throughput sequencing, which has made it easier to sequence organelle genomes from very small amounts of starting material and has led to the publication of several new plastid and mitochondrial genomes in the last few years (DePriest et al. 2013, Janouškovec et al. 2013, Campbell et al. 2014, Salomaki et al. 2015). Thanks to these studies and ongoing efforts in several laboratories, the knowledge of plastid genomes, their evolutionary dynamics, and role in red algal phylogenetics can be expected to expand drastically in the near future.

Red algae have among the largest plastid genomes of any group of algae, generally exceeding 180,000 nt in length and 240 genes. These plastid genomes are most similar to the genome of the cyanobacterial ancestor of primary plastids in the sense that they are so large and have a rich repertoire of protein-coding genes and tRNAs (Lang and Nedelcu 2012). The only exception is the ghost plastid of the alloparasite *Choreocolax polysiphoniae*, which has lost all photosynthetic pathways and has reduced in size to 90 kbp (Salomaki et al. 2015).

Our aim in this study was to add to the knowledge of plastid genomes in the largest order (Ceramiales) and family (Rhodomelaceae) of red algae by sequencing a specimen of the genus *Laurencia*. The Rhodomelaceae are an incredibly diverse family, in terms of variation in thallus types, ecology, and species diversity (the family accounts for 15% of the species diversity of red algae—http://www.algaebase.org/). Two other plastid genomes were recently characterized in this family, those of the photosynthetic macroalga *Vertebrata lanosa* and its nonphotosynthetic alloparasite *Choreocolax polysiphoniae* (Salomaki et al. 2015). At 167 kbp and 90 kbp, these genomes are among the smallest known in red algae, which would suggest a reduction trend present in the photosynthetic macrophytes of this family and greatly accelerated in lineages that lost photosynthesis. Besides characterizing the structure and gene content of the *Laurencia* plastid genome, we also seek to confirm whether there has been a reduction in plastid genome size in the Rhodomelaceae.

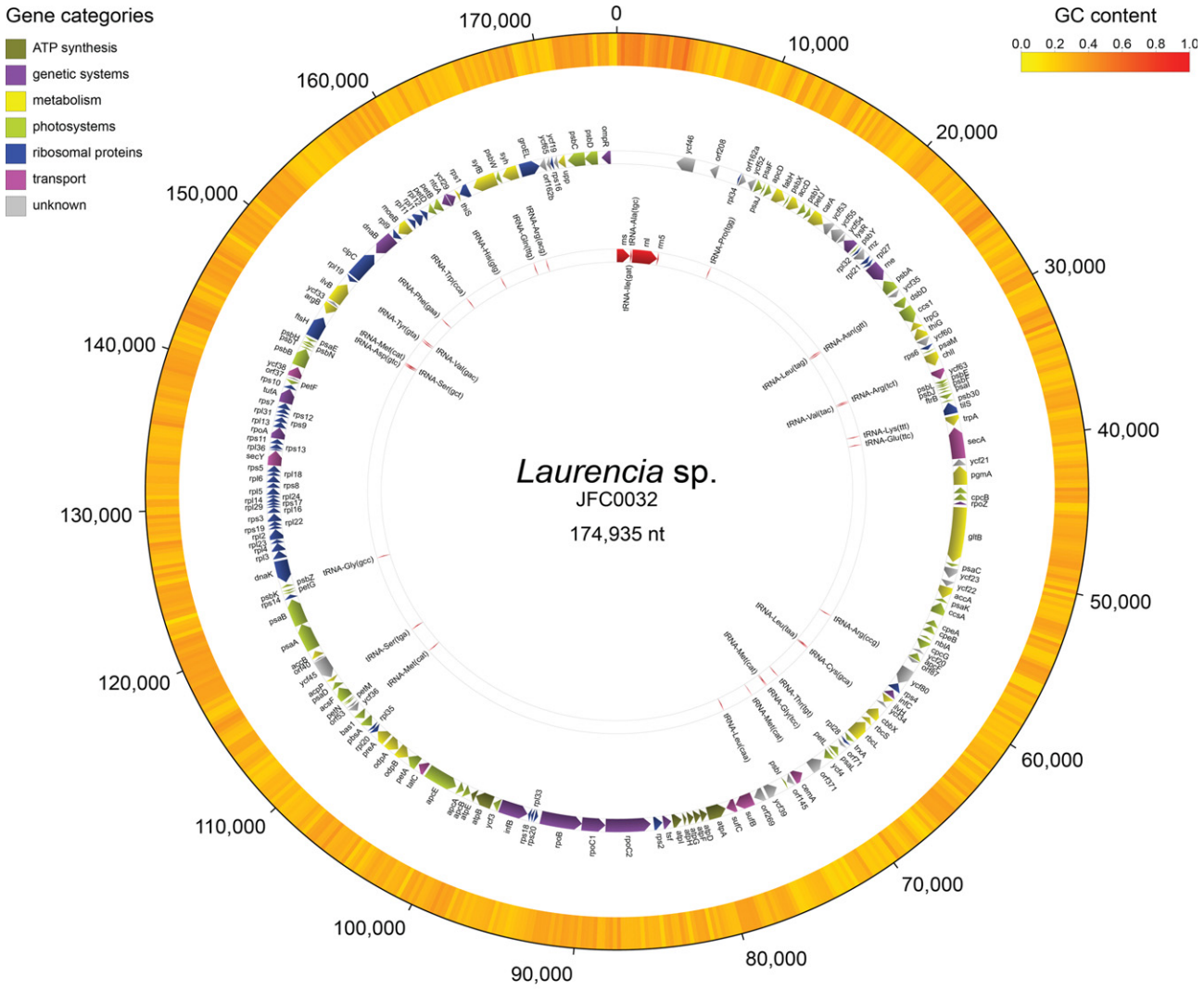
The specimen we sequenced (voucher JFC0032) was collected in Coral Bay, Western Australia, and preserved in silica gel. The sample belongs to an undescribed species of *Laurencia* that is similar in overall morphology to *L. majuscula* and *L. intricata*. An image of the specimen is available on FigShare (http://dx.doi.org/10.6084/m9.figshare.1289784).

Total DNA was isolated with an adapted cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle 1987). In summary, samples were incubated for an hour in CTAB buffer with proteinase K and extracted two times with 24:1 chloroform:isoamyl alcohol. DNA was precipitated using 80% isopropanol at 4°C for 1 h and eluted in 0.1× TE buffer.

Sequencing libraries of the *Laurencia* DNA extract along with 17 other DNA extracts of green and red algae were prepared, each library carrying an individual barcode. Libraries of 350 nt were generated with the TruSeq Nano LT kit and sequenced on an

[1]Received 16 January 2015. Accepted 9 February 2015.
[2]Author for correspondence: e-mail heroen.verbruggen@gmail.com.
Editorial Responsibility: C. Lane (Associate Editor)

A Circular genome map



B Genome alignments

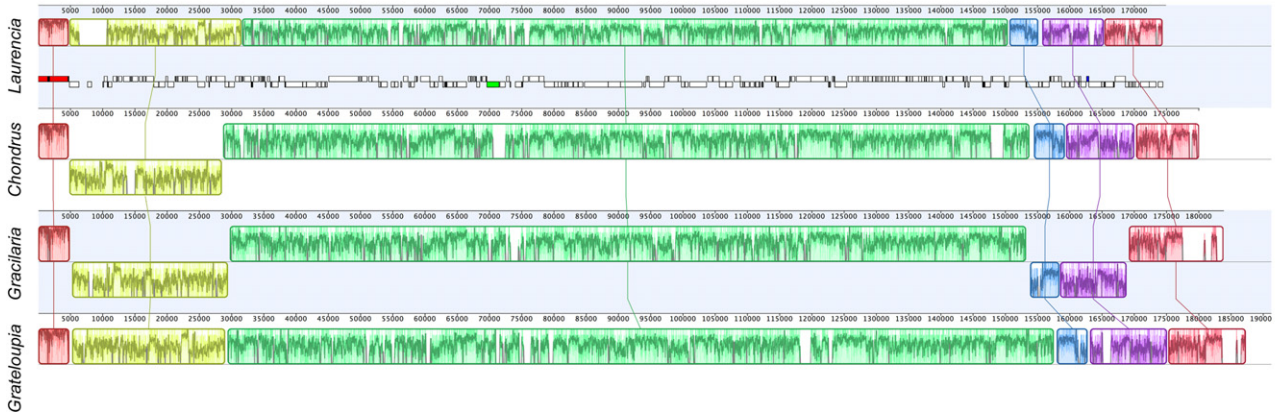


FIG. 1. The plastid genome of *Laurencia* and its alignment with those of other species in the Rhodymeniophycideae. (A) Circular genome map of the *Laurencia* plastid. This image is also available in vector format as well as with CDS coloring indicating protein hydropathy on FigShare (http://dx.doi.org/10.6084/m9.figshare.1289784). (B) Whole-genome alignment of the *Laurencia* plastid genome with that of other sequenced species in the Rhodymeniophycideae, showing large blocks of synteny. The image is also available in vector format from FigShare (same link as above).

Illumina HiSeq 2000 at the Genome Center of the Cold Spring Harbor Marine Laboratory.

The sequencing reads were trimmed with CLC Genomics Workbench 7.5.1 (http://www.clcbio.com), using a quality threshold of 0.05. Sequence assembly was done with MEGAHIT (Li et al. 2015) and CLC Genomics Workbench 7.5.1. For the MEGAHIT assembly, we used 10 kmer sizes (21–91 in steps of 10 and 99) and activated bubble removal. For the CLC assembly, we used automatic word and bubble sizes. One scaffolded area of the plastid genome (84 nt long, in rrn23S) was resolved by mapping the reads to the scaffold in Geneious 8.0.5, using the Geneious mapper at medium–low sensitivity and up to five iterations. The circular-mapping nature of the molecule was verified by mapping the reads against the manually joined end and start of the CLC scaffold.

Annotation used a combination of automated pipelines and manual vetting of the results. The sequence was submitted to the MFannot (http://megasun.bch.umontreal.ca/RNAweasel/), DOGMA (Wyman et al. 2004), and ARAGORN (Laslett and Canback 2004) online tools. DOGMA was run with a 60% cutoff for protein-coding genes, 80% for RNAs, and a BLAST E-value of 1e-5. The resulting annotations were imported into Geneious 8.0.5, manually compared and vetted, and added to the final annotation layer once we were convinced they were correct. The genome was reoriented to have the start of *rrs* (16S) as the first base and submitted to the European Nucleotide Archive (LN833431). Synteny between plastid genomes of *Laurencia* and related species was assessed through whole-genome alignment with the Mauve aligner implemented in Geneious 8.0.5, using automatically calculated seed weights and a fixed minimum LCB score of 30,000. To compare plastid genome sizes in a phylogenetic context, we built a reference phylogeny of the species whose plastid genomes have been fully sequenced. This phylogeny was based on a RAxML analysis of a concatenated 18S + *rbc*L data set with 4 data partitions (18S + 3 codon positions of *rbc*L).

The sequencing run produced 16.8 million 100 nt paired-end reads (2 × 8.4 million) having the barcode corresponding to the *Laurencia* sample. Six percent of these reads (~1 million) assembled into a single scaffold corresponding to the plastid genome, with an average coverage of 590×. MEGAHIT and CLC Genomics Workbench produced scaffolds that were identical except for their start position.

The plastid genome is a 174,935 nt long molecule mapping as a circle (Fig. 1A). It is strongly AT-biased, with only 30% GC on average. The noncoding parts of the genome have very low GC (19%), and protein-coding genes slightly exceed the average with 31% GC. The rRNA cistron forms a block of high GC (46%, Fig. 1A), and the tRNAs are high-GC islands in an AT-rich genome (52% GC on average, Fig. 1). The genome is densely packed (14% noncoding), with spacers between successive protein-coding genes being on average 121 nt long (median is 84 nt), and overlapping ORFs between 6 gene pairs (largest ones are 9 nt between *ccs*1 and *trp*G and 17 nt between *psb*C and *psb*D). These statistics are comparable to those reported for other plastid genomes in the Rhodymeniophycideae. The *Laurencia* genome is, however, among the smallest reported for the florideophytes. A mapping of plastid genome sizes on a phylogenetic tree (Fig. 2) shows that Cyanidiophyceae have slightly smaller genomes, but this is a separate size reduction which could in part be explained by the extreme conditions in which they live (Jain et al. 2014). It seems that in the Rhodomelaceae, an independent reduction has taken place, with the genomes of *Laurencia* and *Vertebrata* being the smallest genomes with the lowest number of protein-coding genes reported to date among photosynthetic florideophytes (Fig. 2). The nonphotosynthetic species *Choreocolax* has seen further reduction of the plastid genome to approximately half the size of those found in typical photosynthetic florideophyte species. A closer look at the

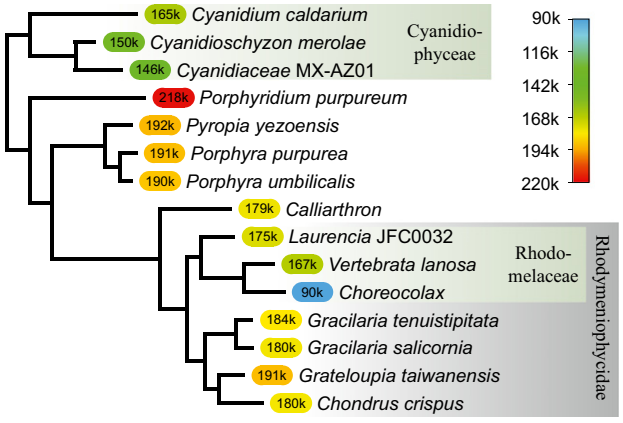


FIG. 2. Phylogenetic mapping of plastid genome sizes in the red algae shows separate size reduction trends in the Cyanidiophyceae and the Rhodomelaceae.

genomes of *Laurencia* and *Vertebrata* shows that the compaction has several reasons, with contributions from the loss of a few genes, somewhat shorter coding sequences (CDS) on average, and slightly shorter intergenic spacers.

The rRNA cistron is present as a single copy and there is no inverted repeat. The genome contains 27 different tRNA genes, with trn-Met(cat) present three times (so 29 tRNA genes in total). Most amino acids have a single tRNA, but Arg and Leu have three each and Gly, Ser, and Val have two each. One of the trn-Met(cat) genes had a 1,876 nt intron containing an unidentified ORF of 1,116 nt. Two additional RNA genes were identified: ARAGORN pointed to a tmRNA, and we also found an RNA region coding for RNase P RNA (*rnp*B). These features are not uncommon in red algal plastid genomes (e.g., Lang and Nedelcu 2012, Janouškovec et al. 2013).

We have annotated 200 protein-coding genes on the *Laurencia* plastid genome, of which 188 are widely conserved and named genes, including the conserved *ycf*# genes. In addition, there were 3 ORFs that showed remote similarity with ORFs from other red algae and 9 hypothetical coding regions that we could not identify based on similarity searches. As is usually the case in florideophyte plastid genomes, no introns were found in any of the protein-coding genes. Most genes use ATG as the start codon (86%) and TAA (69%) or TAG (22%) as the stop codon. The gene content is very similar to that of other florideophyte plastid genomes, with suites of genes encoding ribosomal proteins, the photosystems, ATP synthesis, genetic systems, and basic chloroplast metabolism (Fig. 1A). Only a few genes that are present in most other Rhodymeniophycideae are missing from the *Laurencia* plastid genome (*sec*G, *ycf*17, *ycf*37, *ycf*58, *ycf*80, and *ycf*92), and *Laurencia* plastids do not have any genes that are absent from all other sequenced florideophyte plastids.

Whole-genome alignments show highly conserved architecture among plastid genomes of Rhodymeniophycideae species, with large blocks having complete synteny (Fig. 1B). The red block is the ribosomal cistron, and the remaining blocks consist mostly of protein-coding genes. Synteny of Rhodymeniophycideae with earlier-branching lineages of red algae was studied by DePriest et al. (2013) and Janouškovec et al. (2013). Our genome alignment did not include the photosynthetic species *Vertebrata lanosa*, which also aligns very well with the other Rhodymeniophycideae and the parasite *Choreocolax*, which shares regions of synteny with other Rhodymeniophycideae but has experienced many rearrangements and losses (Salomaki et al. 2015).

In conclusion, *Laurencia* has a smaller plastid genome than most red algae, which taken together with the recently published plastid genome of *Vertebrata* suggests a reduction trend in the Rhodomelaceae that results from a combination of factors including gene loss, shorter CDSs, and reduced intergenic spacers. In the parasitic *Choreocolax*, this trend accelerated with massive reductions of photosynthetic genes. Except for being shorter, the features of the *Laurencia* plastid match those of typical florideophyte plastid genomes, with high gene content and features such as a single-copy ribosomal RNA cistron, a tmRNA, and the RNase P RNA.

Funding was provided by the Australian Research Council (FT110100585 to HV), the Australian Biological Resources Study (RFL213-08 to HV), and the University of Melbourne (MIRS/MIFRS to JFC). We thank Stephanie Muller and Eric Antoniou at the CSHL genome center for sequencing the sample. We are grateful to Yola Metti for comments about the identity of the sample. We thank Mike Van Keulen and Frazer McGregor (Coral Bay Research Station, Murdoch University) and Peter Barnes from Parks WA for facilitating field work.

Campbell, M. A., Presting, G., Bennett, M. S. & Sherwood, A. R. 2014. Highly conserved organellar genomes in the Gracilariales as inferred using new data from the Hawaiian invasive alga *Gracilaria salicornia* (Rhodophyta). *Phycologia* 53:109–16.

DePriest, M. S., Bhattacharya, D. & López-Bautista, J. M. 2013. The plastid genome of the red macroalga *Grateloupia taiwanensis* (Halymeniaceae). *PLoS ONE* 8:e68246.

Doyle, J. J. & Doyle, J. L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 1:11–5.

Jain, K., Krause, K., Grewe, F., Nelson, G. F., Weber, A. P. M., Christensen, A. C. & Mower, J. P. 2014. Extreme features of the *Galdieria sulphuraria* organellar genomes: a consequence of polyextremophily? *Genome Biol. Evol.* 7:367–80.

Janouškovec, J., Liu, S. L., Martone, P. T., Carré, W., Leblanc, C., Collén, J. & Keeling, P. J. 2013. Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. *PLoS ONE* 8:e59001.

Lang, B. F. & Nedelcu, A. M. 2012. Plastid genomes of algae. *In* Bock, R. & Knoop, V. [Eds.] *Genomics of Chloroplasts and Mitochondria*. Springer, New York, New York, pp. 59–87.

Laslett, D. & Canback, B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32:11–6.

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*. doi:10.1093/bioinformatics/btv033.

Salomaki, E. D., Nickles, K. R. & Lane, C. E. 2015. The ghost plastid of *Choreocolax polysiphoniae*. *J. Phycol.* 51:217–21. doi:10.1111/jpy.12283.

Wyman, S. K., Jansen, R. K. & Boore, J. L. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–5.