

# Independent Size Expansions and Intron Proliferation in Red Algal Plastid and Mitochondrial Genomes

Fabian van Beveren , Laura Eme, Purificación López-García , Maria Ciobanu, and David Moreira \*

Ecologie Systématique Evolution, Centre National de la Recherche Scientifique—CNRS, Université Paris-Saclay, AgroParisTech, Orsay, France

\*Corresponding author: E-mail: david.moreira@universite-paris-saclay.fr.

Accepted: 07 March 2022

## Abstract

Proliferation of selfish genetic elements has led to significant genome size expansion in plastid and mitochondrial genomes of various eukaryotic lineages. Within the red algae, such expansion events are only known in the plastid genomes of the Proteorhodophytina, a highly diverse group of mesophilic microalgae. By contrast, they have never been described in the much understudied red algal mitochondrial genomes. Therefore, it remains unclear how widespread such organellar genome expansion events are in this eukaryotic phylum. Here, we describe new mitochondrial and plastid genomes from 25 red algal species, thereby substantially expanding the amount of organellar sequence data available, especially for Proteorhodophytina, and show that genome expansions are common in this group. We confirm that large plastid genomes are limited to the classes Rhodellophyceae and Porphyridiophyceae, which, in part, are caused by lineage-specific expansion events. Independently expanded mitochondrial genomes—up to three times larger than typical red algal mitogenomes—occur across Proteorhodophytina classes and a large shift toward high GC content occurred in the Styloematophyceae. Although intron proliferation is the main cause of plastid and mitochondrial genome expansion in red algae, we do not observe recent intron transfer between different organelles. Phylogenomic analyses of mitochondrial and plastid genes from our expanded taxon sampling yielded well-resolved phylogenies of red algae with strong support for the monophyly of Proteorhodophytina. Our work shows that organellar genomes followed different evolutionary dynamics across red algal lineages.

**Key words:** Proteorhodophytina, Rhodophyta, phylogenomics, group II introns, genome expansion.

## Significance

Red algal plastids exhibit large genome size variation, especially in the recently described subphylum Proteorhodophytina. By contrast, their mitochondrial genomes remain unexplored. Here, we sequenced, assembled, and analyzed new plastid and mitochondrial genomes of 25 species of red microalgae. We uncovered new cases of genome size expansion in plastids and mitochondria, in both cases associated with group II intron proliferation. The phylogenomic analysis supported that these expansions occurred several times independently.

## Introduction

Red algae (Rhodophyta) are an ancient and diverse group of photosynthetic eukaryotes, consisting of more than 7,000 described species (Guiry and Guiry 2021), which range from unicellular species to complex multicellular seaweeds such as nori (*Pyropia yezoensis*) and Irish moss (*Chondrus*

*crispus*). They are characterized by their red plastid pigments and the lack of both flagella and centrioles (Woelkerling 1990). Together with glaucophytes and green algae and plants, they form the supergroup Archaeplastida (Adl et al. 2005). The monophyly of red algae is well supported and the study of their phylogeny has elucidated

© The Author(s) 2022. 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-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

many of their intragroup relationships. The subphylum Eurhodophytina comprises the Bangiophyceae and Florideophyceae, the classes containing seaweeds, whereas the unicellular and extremophilic Cyanidiophyceae solely form the subphylum Cyanidiophytina, sister to the rest of red algae (Qiu et al. 2016). The placement of the remaining four classes (Porphyridiophyceae, Compsopogonophyceae, Rhodellophyceae, and Stylonematophyceae) was long unclear before a phylogenetic study including a representative taxon sampling showed that they were monophyletic and proposed the subphylum Proteorhodophytina to host them (Muñoz-Gómez et al. 2017).

Contemporary plastids of the three phyla of Archaeplastida evolved from an endosymbiotic cyanobacterium through significant metabolic and genomic reduction (Schwartz and Dayhoff 1978; Martin et al. 1998). Among them, the plastid genomes (plastomes) of red algae are usually thought to retain several primitive features and to evolve slowly (Butterfield 2000; Glöckner et al. 2000; Janouškovec et al. 2013). For example, plastome alignments within the classes Bangiophyceae and Florideophyceae show little change in synteny (Janouškovec et al. 2013; Cao et al. 2018), although Cyanidiaceae and inter-class comparisons show less conservation (Janouškovec et al. 2013; Muñoz-Gómez et al. 2017). Genome size shows little variation as well, typically ranging from 150 to 200 kb. By contrast, intron invasion has resulted in massive plastome expansion in the Proteorhodophytina: the Rhodellophyceae include the largest red algal plastid genome (1.1 Mb in *Corynolastis japonica*), more than five times larger than the typical florideophyte plastomes (Muñoz-Gómez et al. 2017).

Self-splicing introns are found not only in plastomes of red algae, but also in the organellar genomes of land plants, green algae, euglenoids, and other protists (Plant and Gray 1988; Copertino and Hallick 1993; Lambowitz and Zimmerly 2011; Brouard et al. 2016). As observed in the Proteorhodophytina, in some of these organisms, such as certain green algae and the euglenoid *Eutreptiella pomquetensis*, the proliferation of self-splicing introns has led to plastid genome expansion (Brouard et al. 2016; Dabbagh et al. 2017). These introns do not only spread within one genome, but they can also travel between organellar genomes, intra- and interspecifically, and even between prokaryotes and eukaryotic organelles (Burger et al. 1999; Sheveleva and Hallick 2004; Pombert et al. 2005; Khan and Archibald 2008). Intron proliferation happens due to the copy-paste ability of the spliced intron, which is aided by an intron-encoded protein (IEP), a maturase often containing a reverse-transcriptase domain (Jacquier and Dujon 1985). However, this copy-paste ability is usually lost in organellar introns, which can have degenerate IEPs like in the plastome of the red alga *Porphyridium purpureum* (Perrineau et al. 2015).

In contrast to their plastomes, genome expansion has never been reported for mitochondrial genomes (mitogenomes) of red algae, which remain much less studied. Although mitogenomes of Florideophyceae species exhibit well-conserved gene content and synteny (Yang et al. 2015), they show more variability in Bangiophyceae and Cyanidiophyceae (Yang et al. 2015; Liu et al. 2020a). Nevertheless, size remains stable within the currently known diversity of red algal mitogenomes, ranging from 21 kb (*Galdieria sulphuraria*) to 43 kb (*Bangia fuscopurpurea*). However, there is only one mitochondrial genome sequenced for the Proteorhodophytina, from the species *Compsopogon caeruleus* (Nan et al. 2017), which hinders the description of the overall mitogenome diversity in all red algal classes.

To fill this gap and better characterize the diversity of organellar genomes in red algae, we have sequenced the plastomes and mitogenomes of 25 species with a particular focus on the poorly known Proteorhodophytina, for which only the *C. caeruleus* mitogenome and few plastomes were available at the start of this work. These new data allowed us to propose a well-resolved phylogenetic framework for Rhodophyta and to unveil divergent evolutionary patterns in these organellar genomes, including multiple genome size expansions.

## Results

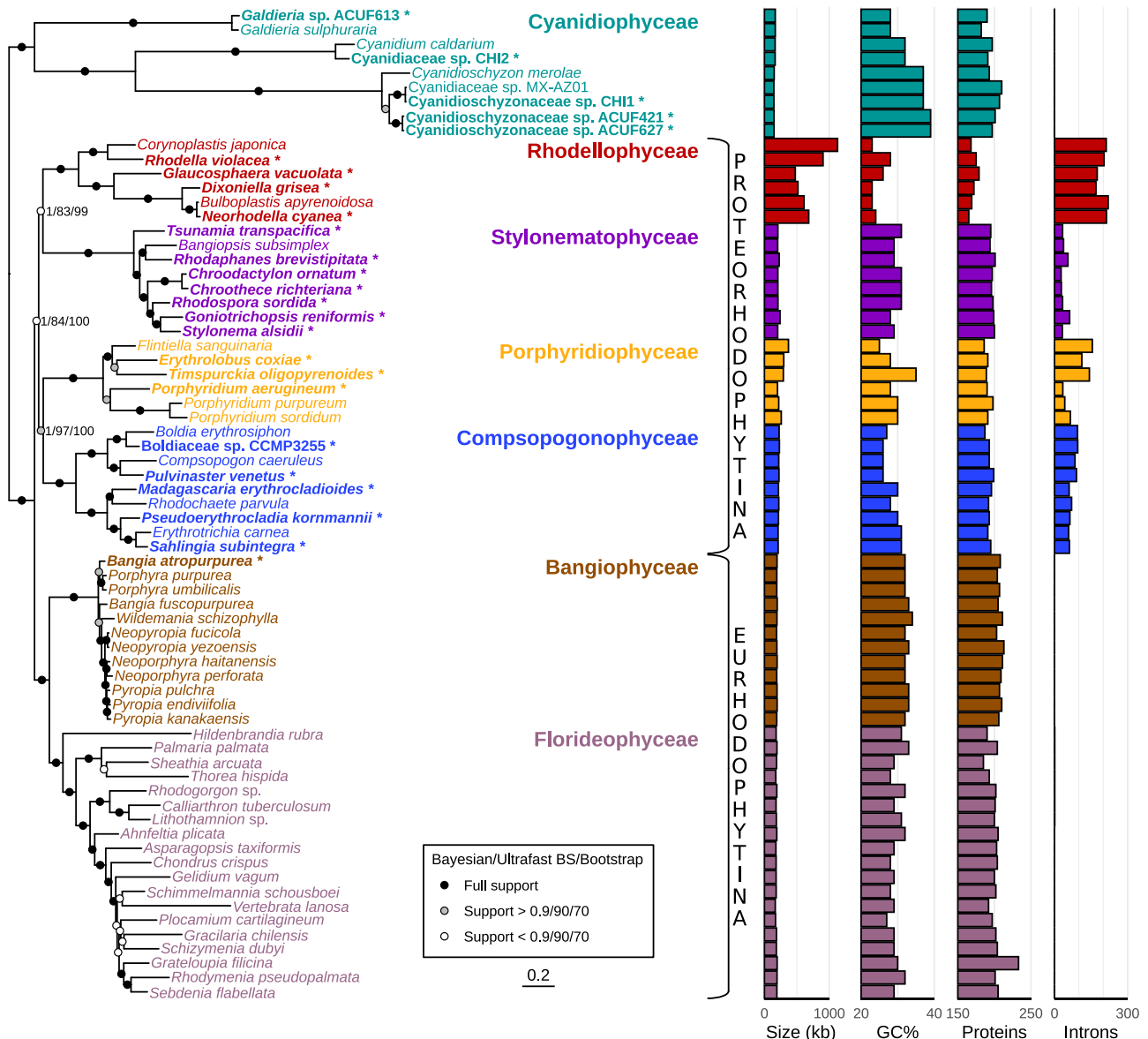
To explore the evolutionary relationships and organellar genome evolution in red algae, we sequenced DNA of 25 red algal species: five Cyanidiophyceae (including two uncultured ones from an environmental sample), seven Stylonematophyceae, three Porphyridiophyceae, five Compsopogonophyceae, four Rhodellophyceae, and one Bangiophyceae ([supplementary table S1, Supplementary Material](#) online; Materials and Methods). We assembled their organellar genome sequences and predicted gene features that we used for genome comparison and phylogenetic analyses.

### Plastid and Mitochondrial Phylogenies Support the Monophyly of Proteorhodophytina

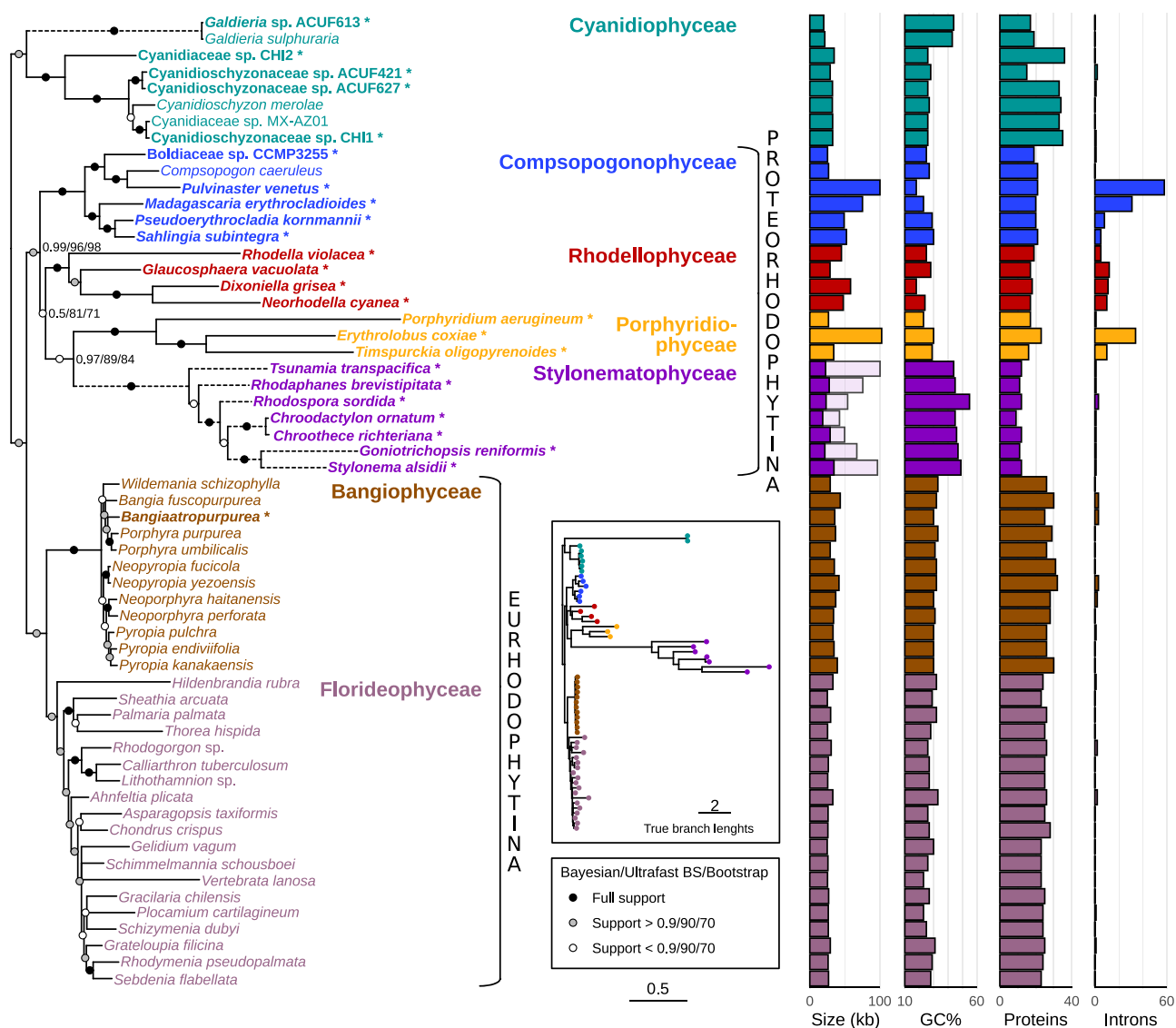
To build a phylogenetic framework for subsequent comparative genomic studies, we first carried out a phylogenomic analysis based on the plastid genomes from this study and those publicly available, including representatives from all known red algal classes. Our plastid data set contained 69 taxa and 189 proteins ([supplementary fig. S1, Supplementary Material](#) online), which, after trimming, consisted of 37,256 conserved amino acid positions. We made a corresponding mitochondrial data set, with 10 fewer taxa (from non-Eurhodophytina lineages lacking available mitochondrial genome sequences). This data set consisted of 59 taxa and 21 proteins (4,537 conserved amino acid positions after trimming; [supplementary fig. S2,](#)

Supplementary Material online). We used Bayesian inference with the CAT + GTR model and maximum likelihood approaches with data set-specific substitution models (see Materials and Methods) to elucidate the evolutionary relationships among red algae with both data sets. These phylogenies were rooted at the base of the Cyanidiophyceae since this class is known to be sister to all other red algae (Ciniglia et al. 2004; Yoon et al. 2006).

Both the plastid and mitochondrial phylogenies recovered the seven known classes of red algae with full support as well as the monophyly of Eurhodophytina (figs. 1 and 2; supplementary fig. S3, Supplementary Material online). The monophyly of Proteorhodophytina was also recovered with strong support for both data sets (plastid: 1/84/100, mitochondrial: 0.99/96/98; support: Bayesian posterior probability/ultrafast bootstrap/nonparametric bootstrap).



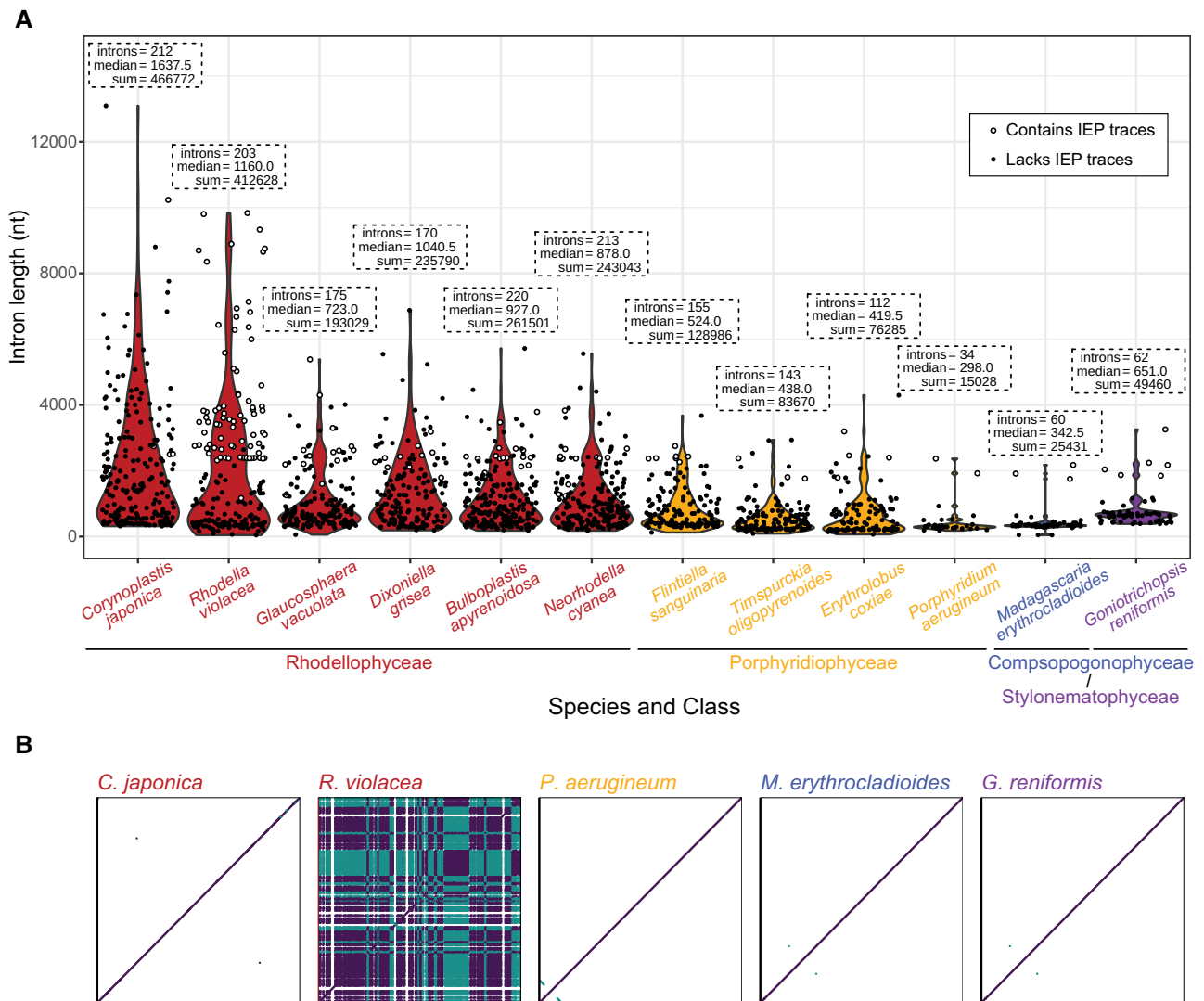
**FIG. 1.**—Phylogenetic tree of red algae based on plastid markers and statistics of plastid genome features. Species names bolded with an asterisk indicate sequences from this study; the colors indicate the different red algal classes. The tree was constructed from a concatenation of 189 proteins (37,256 positions, 69 taxa) using Bayesian inference (CAT + GTR model) and maximum likelihood (cpREV + C60 + F + R7 with 1,000 ultrafast bootstrap replicates; cpREV + C60 + F + R7 + PMSF with 100 nonparametric bootstrap replicates). Support values for the deep relationships among Proteorhodophytina classes are shown with labels (Bayesian probability/ultrafast bootstrap/nonparametric bootstrap). Genome statistics are indicated next to each taxon: genome size (kb), GC content (%), the number of protein-coding genes (notice that the scale ranges between 150 and 250 to emphasize the differences), and the number of introns in protein-coding genes.



**FIG. 2.**—Phylogenetic tree of red algae based on mitochondrial markers and statistics of mitochondrial genome features. Species names with an asterisk indicate sequences from this study; the colors indicate the different red algal classes. The tree was constructed from a concatenation of 21 proteins (4,537 positions, 59 taxa) with Bayesian inference (CAT + GTR) and maximum likelihood (mtZOA + C60 + F + R9 with ultrafast bootstrap; mtZOA + C60 + F + R9 + PMSF with nonparametric bootstrap). Branch support is shown with circles on the branches, actual values are shown for bipartitions important for understanding the relationships among the Proteorhodophytina (Bayesian probability/ultrafast bootstrap/nonparametric bootstrap). All branches for the *Galdieria* clade (*G. phlegrea* + *G. sulphuraria*) and the Stylonematophyceae have been shortened to one-fourth of their actual length for readability (indicated by dashed lines), the true branch lengths are shown in the inset beside the phylogeny. Genome statistics are shown next to each taxon: genome size (kb), GC content (%), the number of protein-coding genes, and the number of introns in protein-coding genes. When mitochondrial genomes were fragmented, a minimal size is shown considering only the contigs with protein-coding genes, and the total size (lighter bar) corresponds to the final assembly including contigs without protein-coding genes.

Additionally, the plastid data set recovered the monophyly of Rhodellophyceae + Stylonematophyceae (1/83/99) as well as the monophyly of Compsopogonophyceae + Porphyridiophyceae (1/97/100) (fig. 1). By contrast, the mitochondrial data set recovered a well-supported clade of Stylonematophyceae + Porphyridiophyceae (0.97/89/84),

but no clear support for any other relationship among the Proteorhodophytina classes (fig. 2). We carried out an approximately unbiased (AU) test (Shimodaira 2002) to compare these two different tree topologies (supplementary fig. S4, Supplementary Material online). Although the topology recovered by the plastid data set (fig. 1) was not



**Fig. 3.**—Introns and repeated regions in a selection of red algal plastid genomes. (A) Violin plots showing the intron size distribution of selected genomes. The total number of introns, the median length of introns, and the total size of all introns are shown in dashed boxes. White dots indicate that the intron has traces of an intron-encoded protein based on BLAST searches of group II intron-encoded reverse transcriptase sequences. (B) Genome self-alignments showing the repetitive content of each genome. Alignments in the forward and reverse directions are colored in purple and cyan, respectively.

rejected by the mitochondrial data set ( $P$ -value = 0.4), the mitochondrial topology (fig. 2) was strongly rejected by the plastid data set ( $P$ -value = 0.00238).

#### Large Plastid Genome Sizes Explained by Relatively Recent Group-II Intron Proliferation

Plastid genomes available so far for the Proteorhodophytina are intron-rich (Tajima et al. 2014; Muñoz-Gómez et al. 2017; Preuss et al. 2021). Our study confirmed this trend; our enriched taxon sampling for this group showed that all currently sequenced Proteorhodophytina plastid genomes had at least 27 introns (fig. 1). Typically, these plastomes contained mostly small (200–600 bp) introns and a

few larger ones (1,500–2,500 bp) that coded for an IEP (fig. 3A). The only exceptions were the highly intron-rich plastomes of Rhodellophyceae and a clade within the Porphyridiophyceae (containing the species *Flintiella sanguinaria*, *Erythrolobus coxiae*, and *Timspurckia oligopyrenoides*), which had introns ranging from small (~100 bp) to huge (up to 13 kb in *Corynoplasis japonica*), with bigger introns often lacking IEPs.

The large number and size of introns in plastomes of Rhodellophyceae and a subset of Porphyridiophyceae explained the large genome sizes in these two clades, ranging from 290 kb up to 1.1 Mb (fig. 1; Muñoz-Gómez et al. 2017). Within the two classes, there was a large variation in genome size. In Porphyridiophyceae, the plastome of

*F. sanguinaria* (370 kb) was significantly larger than the plastomes of its sister lineages *E. coxiæ* (298 kb) and *T. oligopyrenoides* (292 kb), which was at least partly explained by an increase in intron content (fig. 3A). Many differently sized plastomes were present in the Rhodellophyceae, all larger than any of the other red algal plastomes, yet they encoded fewer proteins (fig. 1; supplementary fig. S1, Supplementary Material online). Instead, they contained more and typically larger introns than those found in the expanded plastomes of Porphyridiophyceae (fig. 3A). The two largest red algal plastomes were of the sister lineages *C. japonica* and *Rhodella violacea*, which may suggest that most of the genome expansion occurred in their common ancestor. However, when aligning each genome to itself to find repetitive sequences, there were few highly identical sequences present in *C. japonica*, whereas that of *R. violacea* had many (fig. 3B). These highly identical regions corresponded to introns, many of which contained IEPs, although most were degenerate (fig. 3A; supplementary fig. S5, Supplementary Material online). A possible explanation for this observation is that the plastome of *R. violacea* underwent genome expansion by intron invasion independently of, and more recently than, that of *C. japonica*. An alternative could be that the expansion occurred in the common ancestor of these two species, but that the rate of subsequent intron degeneration was much higher in *C. japonica* than in *R. violacea*.

The analysis of the position of introns can help to distinguish between these two possibilities. If the intron expansion occurred in a common ancestor of *R. violacea* and *C. japonica*, the introns should mostly reside at the same locations, whereas if there were independent intron proliferation events, the introns should often be found at different locations. For all our Proteorhodophytina plastomes, we determined the host gene and the position of each intron and compared it with the other species (fig. 4). Only 42% of the *R. violacea* introns had a corresponding intron at the same position in *C. japonica* (40% for the *C. japonica* versus *R. violacea* comparison). This result, together with the easily detectable IEPs in the plastome of *R. violacea* (supplementary fig. S5, Supplementary Material online) and the close phylogenetic affinity of these IEPs (supplementary fig. S6, Supplementary Material online) supported a single, relatively recent, period of intron proliferation in *R. violacea*. In general, the species of Rhodellophyceae and Porphyridiophyceae had less introns in common with their sister species than the species of Stylonematophyceae and Compsopogonophyceae. Intron proliferation in Rhodellophyceae and Porphyridiophyceae was highly correlated with plastome size (fig. 5A).

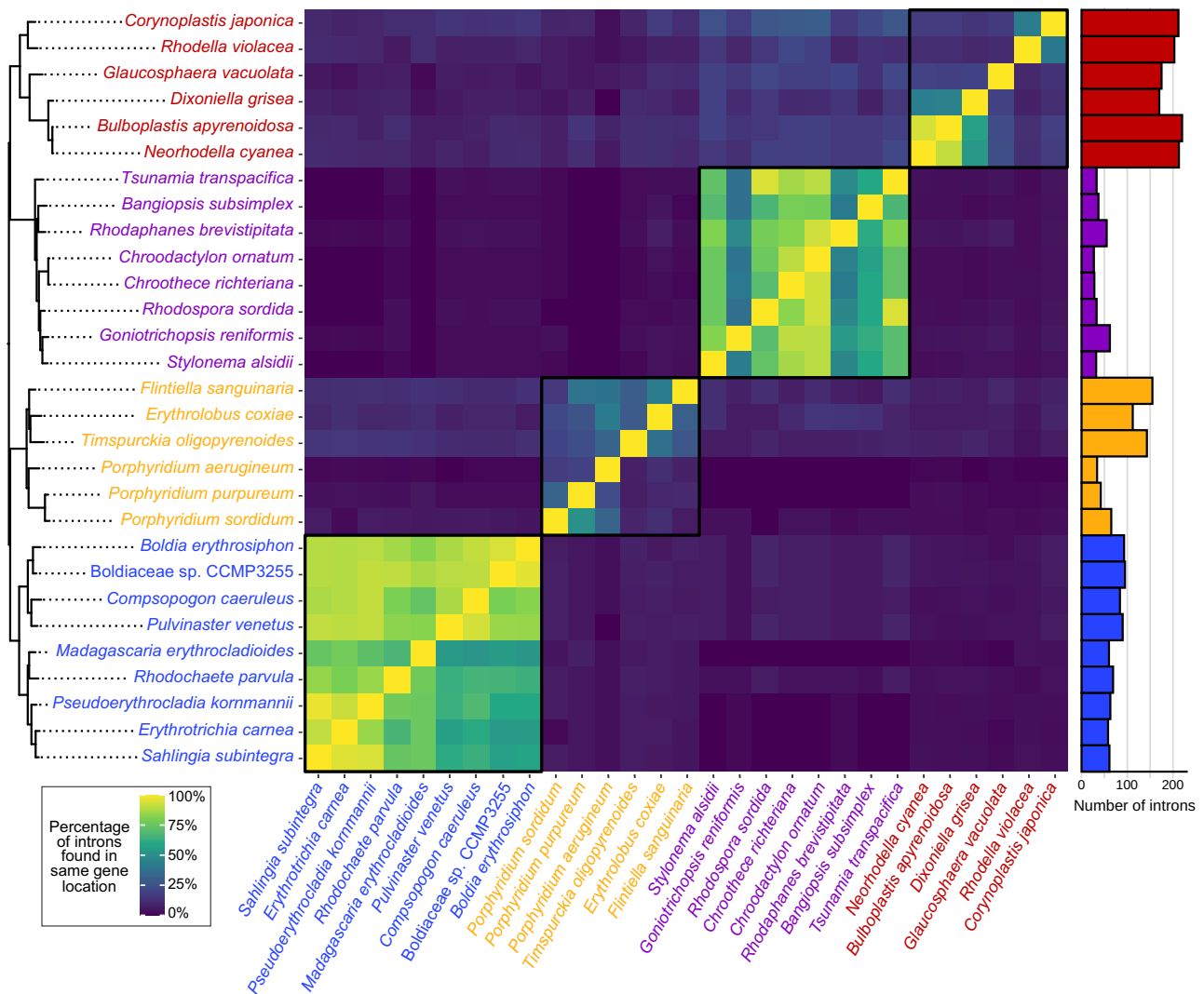
Genome rearrangements are known to have occurred in Proteorhodophytina (Muñoz-Gómez et al. 2017). To detect additional putative recent rearrangements within each class, we aligned the plastomes of their

representatives (supplementary fig. S7, Supplementary Material online). Among plastomes of Stylonematophyceae, only two rearrangements (both inversions) were detected, whereas multiple ones were visible in the plastomes of Compsopogonophyceae, especially between Compsopogonales species (represented here by *Pulvinaster venetus*, *C. caeruleus*, *Boldia erythrosiphon*, and *Boldiaceae* sp. CCMP3255) and the remaining representatives of Compsopogonophyceae. Finally, plastome rearrangements were especially common in representatives of Rhodellophyceae and Porphyridiophyceae, where many occurred among sister lineages. These results suggested different genome rearrangement dynamics in the different red algal classes, potentially caused by the high mobility of introns, especially in the intron-rich plastomes of Rhodellophyceae and Porphyridiophyceae.

### Proteorhodophytina Contain the Largest Mitochondrial Genomes Among Red Algae

All known mitochondrial genomes of red algae are small, ranging from 21 up to 43 kb. This includes *C. caeruleus* (29 kb), the only mitogenome characterized in the Proteorhodophytina (Nan et al. 2017). Like for the plastid genomes, our considerably richer taxon sampling allowed us to reveal a large size variation in red algal mitochondrial genomes (fig. 2), with sizes over twice those previously reported. The largest ones were found in *E. coxiæ* (Porphyridiophyceae, 102 kb) and *P. venetus* (Compsopogonophyceae, 100 kb). The Stylonematophyceae also included highly expanded mitogenomes (e.g., ~100 kb in *Tsunamia transpacificica*), but we only obtained fragmented assemblies for these species, making it difficult to determine their precise size. Overall, although expanded plastid genomes were limited to two specific clades (see above), expanded mitogenomes were found sporadically within different lineages of the Proteorhodophytina (fig. 2).

We examined whether these mitogenome size increases were related to the number of introns. Typically, red algae have up to five mitochondrial introns, including introns in rRNA genes (fig. 2; Yang et al. 2015). We found that the two largest mitogenomes had many more introns interrupting protein-coding genes (58 in *P. venetus* and 34 in *E. coxiæ*) and their presence was correlated with mitogenome size in these species (fig. 5D). However, these intron-rich mitogenomes did not appear as highly rearranged as the intron-rich plastomes of Porphyridiophyceae and Rhodellophyceae representatives (supplementary figs. S7 and S8, Supplementary Material online). Since introns are also common in plastomes, it could be hypothesized that group II introns have been horizontally transferred between the plastid and mitochondrial organellar genomes in different Proteorhodophytina classes. However, a phylogeny of



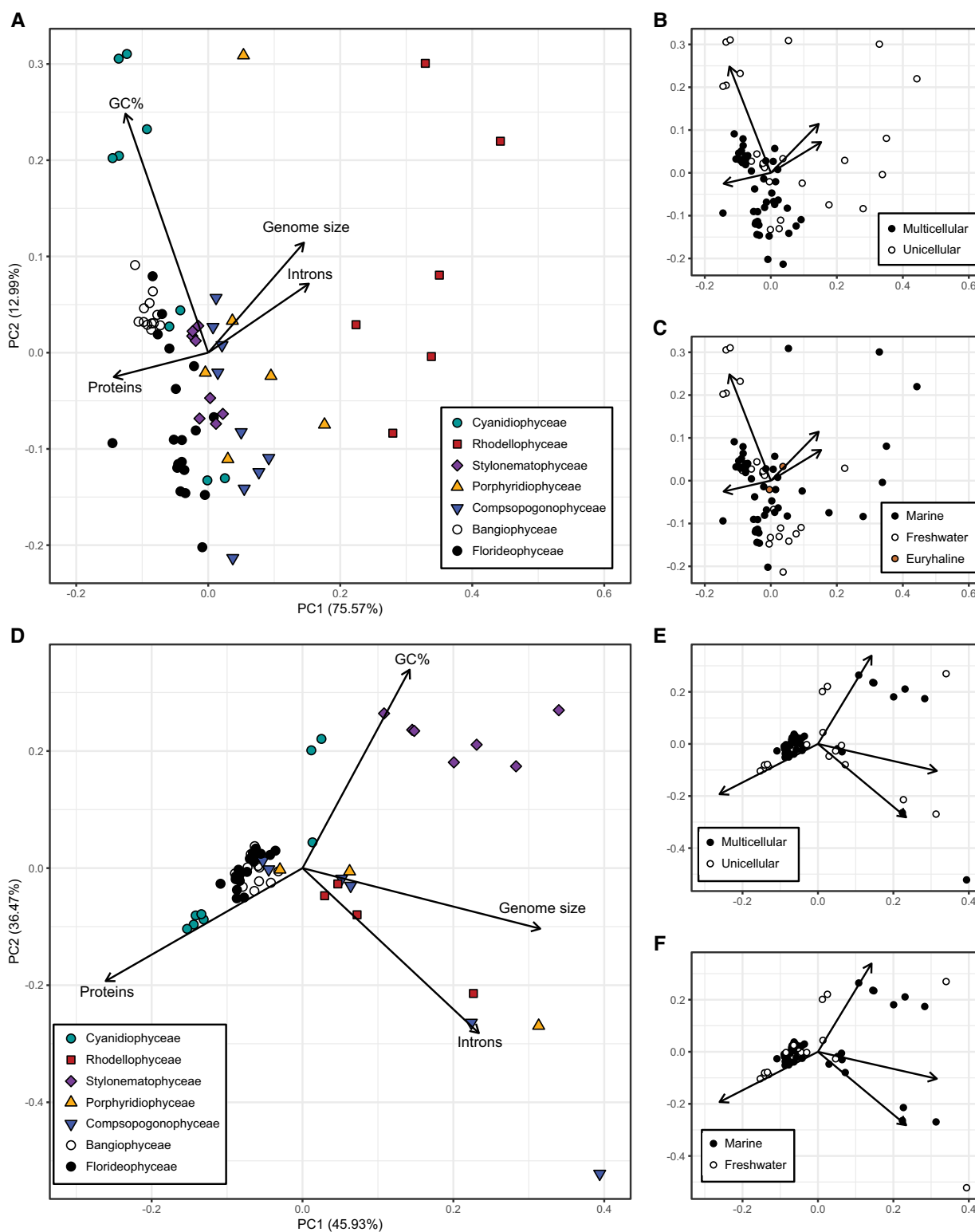
**FIG. 4.**—Percentage of introns present in the same position in red algal plastomes. The position was considered the same if it was in the range of  $\pm 15$  nt in the same gene. Barplots on the right show the total number of introns in each plastome (as in fig. 1).

all red algal intron-encoded proteins did not show any clear recent horizontal transfer of group II introns between the two types of organellar genomes (supplementary fig. S7, Supplementary Material online).

The most divergent mitogenomes were found in the Stylonematophyceae. They contained the smallest number of encoded genes and exhibited the highest GC content (figs. 2 and 5). These mitogenomes also contained only one intron, residing in the *cox3* gene, with the exception of *R. sordida*, which had two more introns. It is possible that the low number of genes and introns identified in these mitogenomes was due to the difficulty to assemble and annotate these more divergent genomes. However, they all appear to lack the same genes, which suggest that our observations are not due to assembly or annotation errors (supplementary fig. S2, Supplementary Material online).

## Discussion

The resolution of the red algal phylogeny, especially the relationships among the various lineages of unicellular and filamentous mesophilic species, has remained controversial for several decades (Gabrielson et al. 1985; Saunders and Hommersand 2004; Yoon et al. 2006). Based on the phylogenomic analysis of six plastid genomes from representatives of different classes, Muñoz-Gómez et al. (2017) observed that these lineages form a monophyletic group, supporting the erection of the subphylum Proteorhodophytina. This result was subsequently confirmed by a phylogenetic analysis of plastid data including additional species of Compsopogonophyceae and Stylonematophyceae (Preuss et al. 2021). In this study, we significantly improved the taxon sampling and showed that the monophyly of Proteorhodophytina is strongly



**Fig. 5.**—PCA of red algal organellar genomes based on genome size, GC content, the number of encoded proteins, and the number of introns. (A–C) PCA for plastid genomes showing (A) the red algal classes to which they belong, (B) the uni- or multicellular character of each species, and (C) the ecosystem type they live in. Euryhaline is used for species known to grow in both freshwater and marine conditions (*Porphyridium purpureum* and *P. sordidum*). (D–F) PCA for mitochondrial genomes showing (D) red algal classes, (E) uni- or multicellular species, and (F) the ecosystem they live in (mitochondrial genome sequences were not available for any euryhaline species).



supported not only based on plastid data, but also on mitochondrial data.

Within Proteorhodophytina, our plastid data set supported the same relationships among classes as found in the previous studies (fig. 1; Muñoz-Gómez et al. 2017; Preuss et al. 2021). The mitochondrial phylogeny showed different relationships among the four Proteorhodophytina classes (fig. 2), but branch support was typically lower and the plastid-derived relationships were not rejected by the mitochondrial data set (supplementary fig. S4, Supplementary Material online). Likely, a combination of (1) the high divergence of the mitochondrial sequences of *Galdieria* sp. ACUF613, *G. sulphuraria*, and Stylonematophyceae (fig. 5D); (2) the substantially smaller size of the mitochondrial data set (4,537 amino acids versus 37,256 for the plastid data set); and (3) the relatively short branches at the base of the Proteorhodophytina resulted in a different, less resolved mitochondrial phylogeny. Altogether this suggests that the plastid-encoded protein data set is more reliable to reconstruct the phylogeny of red algae. Future analyses based on nuclear genomes, still very poorly represented for Proteorhodophytina, will provide an additional test for the robustness of the plastid phylogeny.

Plastid genomes of Cyanidiophyceae, Bangiophyceae, and Florideophyceae are consistent in size and gene density (Janouškovec et al. 2013; Cao et al. 2018; Liu et al. 2020a). By contrast, those of Proteorhodophytina are intron-rich and can reach much larger sizes (fig. 1; Muñoz-Gómez et al. 2017; Preuss et al. 2021). Our larger taxon sampling of Proteorhodophytina allowed us to observe that expanded genomes were restricted to Rhodellophyceae and a subclade within Porphyridiophyceae. As both Rhodellophyceae and Porphyridiophyceae have only unicellular representatives, plastome expansion is limited to unicellular species but includes both freshwater and marine species (fig. 5A–C). A previous comparison of several *P. purpureum* strains revealed that some introns are mobile (Perrineau et al. 2015). Our results supported that this appears to be a general feature in Porphyridiophyceae and Rhodellophyceae (fig. 4), causing plastome expansion in these two red algal classes (fig. 1).

Genome rearrangements are common in the expanded plastomes (supplementary fig. S7, Supplementary Material online). Such rearrangements are often related to the presence of repetitive sequences, for example, in the plastomes of the green alga *Volvox carteri* and many land plants (Smith and Lee 2009; Wicke et al. 2011). Introns have been shown to be directly responsible for genome rearrangements in the bacterial endosymbiont *Wolbachia* (Leclercq et al. 2011). High mobility of group II introns in the plastomes of Rhodellophyceae and Porphyridiophyceae may be the cause of the observed rearrangements, although we cannot exclude that differences in plastome DNA repair

mechanisms may have also played a role (Robart and Zimmerly 2005; Smith 2020). Additional research on the structure of introns and genome comparisons of closely related species will help to better understand the relationship between high intron content and the frequency of genome rearrangement as well as the mechanisms responsible for intron proliferation.

Although expanded genomes in red algae were restricted to the Proteorhodophytina, there are many other algae and land plants known to have large plastid genomes. This is the case of many species of Chlamydomonadales, among which *Haematococcus lacustris* displays the largest known plastome (1.35 Mb) (Bauman et al. 2018; Smith 2018). In contrast with the Proteorhodophytina, these genomes are often expanded due to the proliferation of palindromic repeats, which are also found in mitogenomes of *Volvox* spp. (Aono et al. 2002; Smith and Lee 2009; Zhang et al. 2019).

There is a large variation in mitochondrial genome size among eukaryotes, being famously small and gene-dense in animals and often large in plants—such as the largest known mitogenome of the Siberian larch (*Larix sibirica*) of ~11.7 Mb, which is rich in mobile genetic elements, including group II introns (Putintseva et al. 2020). The mitogenome of the fungus *Morchella crassipes* (~500 kb) is the largest known outside plants and is also rich in group II (but also group I) introns (Liu et al. 2020b). Mitogenome expansion occurs in red algae to a less extreme extent but has resulted in the largest red algal mitochondrial genome (~132 kb), recently described in a strain of *P. purpureum* (Kim et al. 2022). In this study, we show multiple lineage-specific genome expansions, including a more than 3-fold genome size increase in *P. venetus* in comparison to *C. caeruleus* (fig. 2). In contrast to plastome expansions, mitogenome expansions occur not only in unicellular, but also in multicellular (filamentous) species, including both marine and freshwater species (figs. 5D–F).

Organellar introns are known to be mobile, and putative transfers have been reported between red and brown algal mitochondria (Bhattacharya et al. 2001) and from cyanobacteria to red algal mitochondria (Burger et al. 1999). Moreover, intraspecific transfer of intronic elements between organelles was suggested to have occurred in green algae (Pombert et al. 2005; Zhang et al. 2019). One hypothesis is that genome expansion of plastid and mitochondria both result from a general invasion of these organelles by group II introns and/or from intron transfer between them. However, our data do not support these scenarios in red algae since (1) there is only one red algal species (*E. coxiiae*) where both organellar genomes are substantially expanded, and (2) there is no phylogenetic evidence of recent intron transfer between plastid and mitochondrial genomes in the same species based on phylogenies from intron-encoded proteins (supplementary fig. S6,

Supplementary Material online). Although intron transfer between organellar genomes appears not to be responsible for the genome expansions described here, it is clear that selfish genetic elements, in particular group II introns, are largely responsible for organellar genome expansion across the eukaryotic domain, including red algae.

## Materials and Methods

### Species Selection and Culturing

Red algal species were selected based on their phylogenetic position in 18S rRNA and RbCL phylogenies and availability in culture collections. Selected cultures were obtained from the Sammlung von Algenkulturen der Universität Göttingen (SAG; Germany), the Culture Collection of Algae and Protozoa (CCAP; Scotland), the Algal Collection University Federico II (ACUF, Italy), and the National Center for Marine Algae and Microbiota (NCMA; Maine, USA). In addition, we also processed a natural sample rich in cyanidiphyte algae that was collected in the El Chichon volcano (Mexico), which contained two distinct *Galdieria* species. A complete overview of the acquired cultures and method of DNA extraction can be found in [supplementary table S1, Supplementary Material](#) online. All cultures were grown in the laboratory using the Provasoli culture medium for several weeks at 21 °C with a 12 h/12 h light/dark cycle (West and McBride 1999).

### DNA Extraction and Sequencing

Cells were first disrupted by heat-shock, freezing them in liquid nitrogen (90 s), and then thawing them at 65 °C (90 s) three times. Then, different DNA extraction methods were tested to maximize the DNA yield and quality ([supplementary table S1, Supplementary Material](#) online). For all species, the DNeasy PowerBiofilm kit (Qiagen) was used to extract DNA for short-read (Illumina) sequencing, following the manufacturer protocol. Because of a low DNA yield, whole genome DNA amplification was used for *Galdieria phlegrea* and *G. maxima* using the Phi 29 isothermal amplification method with the kit EquiPhi29™ DNA Polymerase (ThermoFisher) as described by the manufacturer. To obtain high molecular weight DNA for long-read (Nanopore) sequencing, we used the DNeasy PowerBiofilm kit (Qiagen) for *R. violacea* and a CTAB-based method for *Sahlinga subintegra*. Briefly, 50 ml of culture was pelleted for 5 min at 500 × g. To eliminate a maximum of polysaccharides and bacteria attached to the red algal cells, the pellet was washed with a mix of culture medium and a nonionic surfactant (Pluronic) at a final concentration of 0.05% by performing three cycles of vortexing, sonication at room temperature (two cycles of 1 min, 37 Hz), and centrifugation (500 g, 5 min). The cleaned pellet was then lysed with 500 µl of Carlston buffer

containing 100 mM Tris-Cl, pH 9.5, 2% CTAB, 1.4 M NaCl, 1% PEG 8000, and 20 mM EDTA, preheated to 65 °C and 300 µl of chloroform, and incubated for 30 min at 65 °C while shaking. The aqueous phase containing the DNA was then further purified with one volume of chloroform/isoamyl alcohol. The DNA was precipitated by incubation with 0.8 M sodium citrate and 1.2 M NaCl and 100% isopropanol at room temperature for 15 min and centrifugation 15 min at 10,000 × g at room temperature. The DNA pellet was washed twice with cold 70% ethanol and dried at room temperature. The DNA was resuspended in 50 µl of preheated (50 °C) 10 mM Tris-HCl pH 8 buffer and quantified with the Qubit™ dsDNA HS Assay Kit (ThermoFisher) following the manufacturer protocol. DNA fragment sizes were visualized on a 0.7% agarose gel stained with GelRed® Nucleic Acid Gel Stain (Biotium). The presence of carryover contaminants was assessed with a Nanodrop spectrophotometer (ThermoFisher).

2 × 150 bp or 2 × 100 bp paired-end Illumina reads were obtained by Eurofins Genomics (Konstanz, Germany) and CNAG-CRG (Barcelona, Spain) using the Illumina HiSeq 2500 and NovaSeq 6000 technologies ([supplementary table S1, Supplementary Material](#) online). Samples prepared for Nanopore sequencing were sequenced on a MinION Mk 1B device on R9.4.1 flow cells. Prior to Nanopore library preparation, the high molecular weight DNA was size-selected with the Short Read Eliminator XS kit (Circulomics) as described by the manufacturer. Nanopore libraries were constructed with the SQK-LSK 109 kit following the Genomic DNA ligation protocol proposed by the manufacturer, with minor modifications: The ligation time was extended to 30 min, all the AmPure beads purification steps were extended to 10 min of incubation with the magnetic beads and the elution from the beads was performed at 37 °C min for several hours. Real time base-calling was performed on a MinIT (MinKNOW v3.6.3) using Guppy (v3.2.9).

### Organellar Genome Assembly

Illumina reads were trimmed with trimmomatic (v0.38; ILLUMINACLIP: adapters.fa: 2: 30: 10 LEADING: 30 TRAILING: 30 SLIDINGWINDOW: 4: 30 MINLEN: 36; Bolger et al. 2014) using the extended list of adapters from BBMap (v38.41; Bushnell 2014). The quality of trimmed Illumina reads was checked with FastQC (v0.11.5; Andrews 2016). Genomes sequenced with both Nanopore and Illumina were assembled with Unicycler (v0.4.9b; using bold mode for *R. violacea* and conservative mode for *S. subintegra*; Wick et al. 2017). The plastid genomes of *Neorhodella cyanea* and *Rhodophanes brevistipitata* were assembled with GetOrganelle (v1.5.1c; with variable k-mers and -R 200; Jin et al. 2020) due to the presence of many short repeats and a single large repeated

region, respectively. All other plastid genomes were assembled using NOVOPlasty (v3.4; Dierckx et al. 2017), with the *rbcl* sequence in the assembly as the seed sequence and variable k-mer sizes. No single assembler worked optimally for all mitogenomes and NOVOPlasty, GetOrganelle, and SPAdes (v3.11.0; with or without meta; Nurk et al. 2017; Prjibelski et al. 2020) were used (see [supplementary table S1, Supplementary Material](#) online). Genome assemblies obtained with GetOrganelle and SPAdes were manually inspected using Bandage (Wick et al. 2015). When an organellar genome assembly remained fragmented, only the contigs with protein-coding genes of plastid or mitochondrial origin were selected. For the mitogenomes of the Stylonematophyceae, this resulted in a large discrepancy between the assembled genome size and the cumulative size of selected contigs; thus, both were used to estimate genome size and other genome statistics (fig. 2).

Organellar reads were gathered by mapping the reads to the assembled organellar contigs, using BBmap with option paired only = t for short reads, and Minimap2 (v2.17-r941; Li 2018) with option -ax map-ont followed by *samtools fastq* (v1.9-52-g651bf14; Li et al. 2009) with option -F 4 for long reads. All organellar reads and unfragmented organellar genomes are available on SRA (BioProject PRJNA744153) and GenBank, respectively ([supplementary table S1, Supplementary Material](#) online). Fragmented organellar genomes are provided as fasta files, along with the predicted protein sequences (see fragmented organellar genome data in the Figshare repository).

To identify plastome and mitogenome rearrangements within non-Eurhodophytina classes, we aligned the organellar genomes of representatives for each class ([supplementary figs. S7 and S8, Supplementary Material](#) online). As starting points of circular genomes are arbitrary, the genomes were aligned with Mauve (snapshot 2015-02-13; Darling et al. 2004) to identify a common syntenic region to use as the starting point. The adjusted plastomes were then aligned using Mauve with default settings for each of the classes.

### Genome Annotation

tRNAs were predicted using TRNAscan-SE with parameter -O (v2.0.3; Chan and Lowe 2019) and rRNAs were predicted using rnammer with parameters -S bac -m lsu, ssu, tsu (v1.2; Lagesen et al. 2007). Protein-coding sequences were predicted using protein sequences of closely related species with published organellar genomes as reference. Exonerate (v2.3; Slater and Birney 2005) was used to align known proteins to the genome assemblies as it can infer introns, using the applicable genetic code and a variation of allowed intron sizes (other options: -model protein2dna: bestfit -E -n 1 -s 60 -percent 10). Nucleotides for splice sites

were assumed unknown, as manual inspection of the protein sequences did not show consistent patterns (data not shown). An in-house python script (*cds\_from\_exonerate.py*, available in Figshare) was used to create a valid coding sequence based on the Exonerate data by finding a correct start codon near the start of the protein alignment and a stop codon near the end. Manual curation of the genome annotation was done using BLAST (Johnson et al. 2008) and the Artemis genome browser (Carver et al. 2012). The published plastid genome sequence of *C. caeruleus* was reannotated using the same method, as intron predictions were inconsistent with that of other Proteorhodophytina plastomes. The number of predicted introns in plastid-encoded proteins of *C. caeruleus* changed from 20 to 85.

### Genome Statistics

To compare the organellar genomes, genome length and GC content were determined using *stats.sh* from the BBTools suite (Bushnell 2014). For the fragmented mitogenomes of the Stylonematophyceae, it was run both for the complete mitochondrial assembly ("maximum size") and the protein-coding contigs ("minimum size"). The number of encoded proteins was estimated by counting the number of genes represented by CDS features in the .gff files. For each gene, the number of introns was assumed to be equal to the number of CDS features minus one. A principal component analysis (PCA) was made based on these statistics for both the plastid and mitochondrial genomes using the R stats function *prcomp* (R v4.0.4; options scale.=T; R Core Team 2021). PCA plots were made using the ggfortify function *autoplot* (v0.4.11; Tang et al. 2016) in ggplot2 (v3.3.3; Wickham 2016).

### Phylogenetic Analyses

Published sequence data of organellar genomes from other red algal species were retrieved from GenBank ([supplementary table S1, Supplementary Material](#) online; Benson et al. 2013). In total, 43 plastid genomes were collected comprising all species of Cyanidiophyceae and Proteorhodophytina with available plastid genomes as well as a selection of Florideophyceae and Bangiophyceae. Including our own data, this resulted in a data set of 69 taxa (plastid dataset in Figshare). Among these species, ten of the Cyanidiophyceae and Proteorhodophytina representatives had no mitochondrial genome sequence available, which resulted in a mitochondrial data set containing 59 taxa (mitochondrial dataset in Figshare).

For each data set, we used an all-against-all BLAST search using psi-blast (BLAST 2.6.0+; options: -evalue 10 -outfmt 6 -max\_target\_seqs 200 -seg yes -soft\_masking true -use\_sw\_tback -word\_size 2 -matrix BLOSUM45;

Altschul et al. 1997), and draft orthologous groups were created using orthAgogue (v1.0.3; default options; Ekseth et al. 2014) and running mcl (v1: 14-137; van Dongen 2000) on the "all.abc" output file. An in-house script was used to deal with the presence of paralogous sequences (mcl\_to\_og.py, available in Figshare). These were either merged if the separated genes were considered two portions of a unique gene separated due to misannotation, or in the case of duplicated sequences specific to a genome, the sequence most similar to other sequences in the orthologous group was retained.

For the plastid and mitochondrial data sets, the multiple sequence alignment (MSA) for each protein was constructed with MAFFT G-INS-i or MAFFT L-INS-i, respectively (v7.310; Katoh et al. 2005). The MSAs were manually refined using single-protein trees made with IQ-TREE (v1.6.11; using -m cpREV + C60 + F + G for plastid MSAs, and -mset LG -mrate G, I + G, R -mfreq FU, F for mitochondrial MSAs; Nguyen et al. 2015) and inspecting the alignments with Aliview (v1.24; Larsson 2014). Final MSAs were trimmed using BMGE (default options; Criscuolo and Gribaldo 2010) and concatenated, resulting in a plastid data set of 37,256 amino acids and a mitochondrial data set of 4,537 amino acids.

The concatenated plastid data set was analyzed by maximum likelihood (ML) with IQ-TREE using the model cpREV + C60 + F + R7 and ultrafast bootstrap (Nguyen et al. 2015; Hoang et al. 2018). The resulting ML tree was used as the guide tree for rapid approximation of posterior mean site frequency (PMSF) under the same model and the generation of 100 nonparametric bootstrap replicates (Wang et al. 2018). The mitochondrial data set was similarly analyzed with the models mtZOA + C60 + F + R9 and mtZOA + PMSF(C60)+F + R9. The Bayesian inference phylogenetic analysis of the plastid and mitochondrial data sets was performed using Phylobayes-MPI (v1.8c; Lartillot et al. 2013) with the CAT-GTR model (Lartillot and Philippe 2004) and four chains run to 10,000 generations (maxdiff remained 1 for both data sets due to the lack of resolution of several branches within the Florideophyceae). Trace files were inspected with graphylo (supplementary figs. S9 and S10, Supplementary Material online; <https://github.com/wrf/graphphylo>) and Tracer (Rambaut et al. 2018) to determine a burn in of 200 generations for both mitochondrial and plastid phylogenies. Phylogenetic trees were plotted with the R package ggtree (v2.4.1; Yu 2020). The above phylogenies, along with the full alignments and single gene alignments, are available in the Figshare repository.

To compare the trees based on the plastid and mitochondrial data sets, we reconstructed trees for both the plastid and mitochondrial taxon sampling and constrained two possible relationships within the Proteorhodophytina: (1) a clade of Stylonematophyceae + Porphyridiophyceae,

with remaining relationships unresolved, and (2) a clade of Stylonematophyceae + Rhodellophyceae, plus a clade of Compsopogonophyceae + Porphyridiophyceae (supplementary fig. S4, Supplementary Material online). All other relationships were kept the same as in the original trees. For both data sets, we tested the three trees from the original analysis (Bayesian and maximum likelihood with and without PMSF) and 100 nonparametric bootstrap trees of the related maximum likelihood analysis with PMSF. The AU-test (Shimodaira 2002) was used, as implemented in IQ-TREE, with the options -n 0 -zb 10,000 -au -zw (Nguyen et al. 2015). The models used were as before: mtZOA + C60 + F + R9 for the mitochondrial data set cpREV + C60 + F + R7 for the plastid data set.

A single-gene phylogeny using the protein RbcL was made to include more available plastid data for comparison (RbcL data set in Figshare). We gathered RbcL sequences of the 69 representatives in the plastid data set and all RbcL sequences of red algal origin not part of the Eurhodophytina from Uniprot (The UniProt Consortium 2015), resulting in a total of 404 sequences. Sequences were aligned with MAFFT L-INS-i (Katoh et al. 2005) and phylogenetic inference was done using IQ-TREE (Nguyen et al. 2015) with options -m cpREV + C40 + F + R4 -bb 1,000 -alrt 1,000 (supplementary fig. S11, Supplementary Material online).

### Analysis of Introns and Intron-Encoded Proteins

First, for each intron interrupting a protein-coding gene in plastid genomes, we determined to which orthologous group it was associated (based on ortholog grouping, see "Phylogenetic Analyses" above). Second, the intron location within the gene was determined by considering the first nucleotide of the start codon as position 0 and disregarding the size of previous introns within that gene sequence (i.e., as if the introns were spliced out). To estimate the number of plastome introns that a species A shared with another species B, we calculated the fraction of introns of species A that are at the same location ( $\pm 15$  nt) in the same gene in species B. These results were visualized with a heatmap made with ggplot2 (Wickham 2016).

To detect whether plastid introns have traces of IEPs, they were aligned to a set of 144 proteins with the reverse-transcriptase domain of group II introns from the conserved domain database (CDD; cd01651; Lu et al. 2020) using blastx (*E*-value threshold of  $1 \times 10^{-10}$ ; Altschul et al. 1997). Size distribution of the introns of different species was plotted using violin plots and raw data points with ggplot2 (Wickham 2016).

To construct a phylogeny of all IEPs encoded in red algal group II introns, we used the cd01651 protein sequences to search all our plastid and mitochondrial genomes and all red algal organellar genomes in GenBank using tblastn

with the option `-max_intron_length 15,000` (Altschul et al. 1997). Regions with blast hits were considered to have traces of an IEP if there was more than one blast hit and an  $E$ -value  $< 1 \times 10^{-5}$ . These parameters were chosen based on manually checking the results obtained with the plastid genome of *E. coxiæ*. Sequences were aligned using MAFFT L-INS-i (Kato et al. 2005) and nonhomologous sequences were removed. The final alignment consisted of 546 IEP sequences and 2,084 amino acid positions and was trimmed with BMGE (to 462 amino acids) with the options `-w 1 -h 1 -g 0.7` (Crisuolo and Gribaldo 2010). The IEP phylogeny was reconstructed with IQ-TREE (`-m LG + C20 + G + F -bb 1,000`) and with the R package `ggtree` (supplementary fig. S6, Supplementary Material online; intron data set in Figshare; Yu 2020).

To detect the presence of repetitive regions, including multiple copies of intron-encoded proteins and other repeats, we aligned each genome with itself using NUCmer (`-maxmatch -nosimplify`; MUMmer, 4.0.0beta2; Marçais et al. 2018). Files with coordinates were created using `show-coords (-c -l -r -T)` and plots were created using `ggplot2`.

## Supplementary Material

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

## Acknowledgments

We thank the UNICELL single-cell genomics platform (<https://www.deemteam.fr/en/unicell>) for help in DNA preparation and Nanopore sequencing, Iris Rizos for preliminary analysis on intron size distribution in red algal plastids, and Line Le Gall for critical reading of the manuscript. We are grateful for the comments and suggestions of two anonymous reviewers that helped to improve this article. This work was funded by the European Research Council Advanced Grants ProtistWorld (No. 322669, P.L.-G.) and Plast-Evol (No. 787904, D.M.) and the ERC Starting grant MacroEpik (No. 803151, L.E.).

## Data Availability

The Nanopore and Illumina sequencing reads were deposited in the NCBI SRA under BioProject PRJNA744153. Complete organellar genomes were submitted to GenBank, accessions can be found in supplementary table S1, Supplementary Material online. Sequencing reads and organellar genomes are also available at Figshare (<https://doi.org/10.6084/m9.figshare.17111693>), along with fragmented organellar genome and protein sequences, phylogenomic data sets, and custom python scripts.

## Literature Cited

- Adl SM, et al. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol.* 52(5):399–451.
- Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17):3389–3402.
- Andrews S. 2016. FastQC: A quality control tool for high throughput sequence data. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Aono N, Shimizu T, Inoue T, Shiraishi H. 2002. Palindromic repetitive elements in the mitochondrial genome of *Volvox*<sup>1</sup>. *FEBS Lett.* 521(1):95–99.
- R Core Team. 2021. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: <https://www.R-project.org/>.
- Bauman N, et al. 2018. Next-generation sequencing of *Haematococcus lacustris* reveals an extremely large 1.35-megabase chloroplast genome. *Genome Announc.* 6(12):e00181-18.
- Benson DA, et al. 2013. GenBank. *Nucleic Acids Res.* 41(D1):D36–D42.
- Bhattacharya D, Cannone JJ, Gutell RR. 2001. Group I intron lateral transfer between red and brown algal ribosomal RNA. *Curr Genet.* 40(1):82–90.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
- Brouard J-S, Turmel M, Otis C, Lemieux C. 2016. Proliferation of group II introns in the chloroplast genome of the green alga *Oedocladium carolinianum* (Chlorophyceae). *PeerJ.* 4:e2627.
- Burger G, Saint-Louis D, Gray MW, Lang BF. 1999. Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*. Cyanobacterial introns and shared ancestry of red and green algae. *Plant Cell* 11(9):1675–1694.
- Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. Berkeley (CA): Lawrence Berkeley National Lab (LBNL). Available from: <https://www.osti.gov/biblio/1241166-bbmap-fast-accurate-splice-aware-aligner>.
- Butterfield NJ. 2000. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* 26(3):386–404.
- Cao M, Bi G, Mao Y, Li G, Kong F. 2018. The first plastid genome of a filamentous taxon '*Bangia*' sp. OUCPT-01 in the Bangiales. *Sci Rep.* 8:10688.
- Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28(4):464–469.
- Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. In: Kollmar M, editor. *Gene prediction*. *Methods Mol Biol.* Vol. 1962. New York (NY): Humana. pp. 1–14.
- Ciniglia C, Yoon HS, Pollio A, Pinto G, Bhattacharya D. 2004. Hidden biodiversity of the extremophilic Cyanidiales red algae. *Mol Ecol.* 13(7):1827–1838.
- Copertino DW, Hallick RB. 1993. Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. *Trends Biochem Sci.* 18(12):467–471.
- Crisuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol.* 10(1):210.
- Dabbagh N, Bennett MS, Triemer RE, Preisfeld A. 2017. Chloroplast genome expansion by intron multiplication in the basal psychrophilic euglenoid *Eutreptiella pomquetensis*. *PeerJ.* 5:e3725.

- Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14(7):1394–1403.
- Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45(4):e18.
- Ekseth OK, Kuiper M, Mironov V. 2014. orthAgogue: an agile tool for the rapid prediction of orthology relations. *Bioinformatics* 30(5): 734–736.
- Gabrielson PW, Garbary DJ, Scagel RF. 1985. The nature of the ancestral red alga: inferences from a cladistic analysis. *Biosystems* 18(3): 335–346.
- Glöckner G, Rosenthal A, Valentin K. 2000. The structure and gene repertoire of an ancient red algal plastid genome. *J Mol Evol.* 51(4):382–390.
- Guiry MD, Guiry GM. 2021. Algaebase. Galway: AlgaeBase World-wide electronic publication, National University of Ireland. Available from: <https://www.algaebase.org> (Last accessed August 25, 2021).
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 35(2):518–522.
- Jacquier A, Dujon B. 1985. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* 41(2):383–394.
- Janouškovec J, et al. 2013. Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. *PLoS One* 8(3):e59001.
- Jin J-J, et al. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 21(2): 241.
- Johnson M, et al. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36(suppl\_2):W5–9.
- Katoh K, Kuma K, Miyata T, Toh H. 2005. Improvement in the accuracy of multiple sequence alignment program MAFFT. *Genome Inform.* 16(1):22–33.
- Khan H, Archibald JM. 2008. Lateral transfer of introns in the cryptophyte plastid genome. *Nucleic Acids Res.* 36(9):3043–3053.
- Kim D, et al. 2022. Group II intron and repeat-rich red algal mitochondrial genomes demonstrate the dynamic recent history of autocatalytic RNAs. *BMC Biol.* 20(1):2.
- Lagesen K, et al. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35(9):3100–3108.
- Lambowitz AM, Zimmerly S. 2011. Group II Introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect Biol.* 3(8):a003616.
- Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30(22): 3276–3278.
- Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol.* 21(6):1095–1109.
- Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol.* 62(4):611–615.
- Leclercq S, Giraud I, Cordaux R. 2011. Remarkable abundance and evolution of mobile group II introns in *wolbachia* bacterial endosymbionts. *Mol Biol Evol.* 28(1):685–697.
- Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18):3094–3100.
- Liu W, et al. 2020b. Subchromosome-scale nuclear and complete mitochondrial genome characteristics of *Morchella crassipes*. *Int J Mol Sci.* 21(2):483.
- Liu SL, Chiang Y-R, Yoon HS, Fu H-Y. 2020a. Comparative genome analysis reveals *Cyanidiococcus* gen. nov., a new extremophilic red algal genus sister to *Cyanidioschyzon* (Cyanidioschyzonaceae, Rhodophyta). *J Phycol.* 56(6):1428–1442.
- Lu S, et al. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48(D1):D265–D268.
- Marçais G, et al. 2018. MUMmer4: a fast and versatile genome alignment system. *PLOS Comput Biol.* 14(1):e1005944.
- Martin W, et al. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393:162–165.
- Muñoz-Gómez SA, et al. 2017. The new red algal subphylum Proteorhodophytina comprises the largest and most divergent plastid genomes known. *Curr Biol.* 27(11):1677–1684.e4.
- Nan F, et al. 2017. Origin and evolutionary history of freshwater Rhodophyta: further insights based on phylogenomic evidence. *Sci Rep.* 7:2934.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
- Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27(5): 824–834.
- Perrineau M-M, Price DC, Mohr G, Bhattacharya D. 2015. Recent mobility of plastid encoded group II introns and twintrons in five strains of the unicellular red alga *Porphyridium*. *PeerJ.* 3:e1017.
- Plant AL, Gray JC. 1988. Introns in chloroplast protein-coding genes of land plants. *Photosynth Res.* 16(1):23–39.
- Pombert J-F, Otis C, Lemieux C, Turmel M. 2005. The chloroplast genome sequence of the green alga *Pseudoclonium akinetum* (Ulvophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages. *Mol Biol Evol.* 22(9):1903–1918.
- Preuss M, Verbruggen H, West JA, Zuccarello GC. 2021. Divergence times and plastid phylogenomics within the intron-rich order Erythropeliales (Compsopogonophyceae, Rhodophyta). *J Phycol.* 57(3):1035–1044.
- Prijbelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes de novo assembler. *Curr Protoc Bioinformatics* 70(1): e102.
- Putintseva YA, et al. 2020. Siberian larch (*Larix sibirica* Ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. *BMC Genomics* 21(1):654.
- Qiu H, Yoon HS, Bhattacharya D. 2016. Red algal phylogenomics provides a robust framework for inferring evolution of key metabolic pathways. *PLoS Curr.* 8, doi:10.1371/currents.tol.7b037376e6d84a1be34af756a4d90846.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol.* 67(5):901–904.
- Robart AR, Zimmerly S. 2005. Group II intron retroelements: function and diversity. *Cytogenet Genome Res.* 110(1-4):589–597.
- Saunders GW, Hommersand MH. 2004. Assessing red algal supraordinal diversity and taxonomy in the context of contemporary systematic data. *Am J Bot.* 91(10):1494–1507.
- Schwartz RM, Dayhoff MO. 1978. Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science* 199(4327):395–403.
- Sheveleva EV, Hallick RB. 2004. Recent horizontal intron transfer to a chloroplast genome. *Nucleic Acids Res.* 32(2):803–810.
- Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst Biol.* 51(3):492–508.
- Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6(1):31.

- Smith DR. 2018. *Haematococcus lacustris*: the makings of a giant-sized chloroplast genome. *AoB PLANTS* 10(5):ply058.
- Smith DR. 2020. Can green algal plastid genome size be explained by DNA repair mechanisms? *Genome Biol Evol.* 12(2):3797–3802.
- Smith DR, Lee RW. 2009. The mitochondrial and plastid genomes of *Volvox carteri*: bloated molecules rich in repetitive DNA. *BMC Genomics* 10:132.
- Tajima N, et al. 2014. Analysis of the complete plastid genome of the unicellular red alga *Porphyridium purpureum*. *J Plant Res.* 127(3): 389–397.
- Tang Y, Horikoshi M, Li W. 2016. ggfortify: unified interface to visualize statistical results of popular R packages. *R J.* 8(2):474.
- The UniProt Consortium. 2015. UniProt: a hub for protein information. *Nucleic Acids Res.* 43(D1):D204–D212.
- van Dongen SM. 2000. Graph clustering by flow simulation [Doctoral dissertation]. [Utrecht (Netherlands)]: University Utrecht.
- Wang H-C, Minh BQ, Susko E, Roger AJ. 2018. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst Biol.* 67(2): 216–235.
- West JA, McBride DL 1999. Long-term and diurnal carpospore discharge patterns in the Ceramiaceae Rhodomelaceae and Delesseriaceae (Rhodophyta). *Hydrobiol.* 398-399:101–114.
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput Biol.* 13(6):e1005595.
- Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31(20): 3350–3352.
- Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.* 76(3): 273–297.
- Wickham H. 2016. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag. Available from: <https://ggplot2.tidyverse.org>.
- Woelkerling WMJ. 1990. An Introduction. In: Cole KM, Sheath RG, editors. *Biology of the Red Algae*. New York (NY): Cambridge University Press. pp. 1–6.
- Yang EC, et al. 2015. Highly conserved mitochondrial genomes among multicellular red algae of the Florideophyceae. *Genome Biol Evol.* 7(8):2394–2406.
- Yoon HS, Muller KM, Sheath RG, Ott FD, Bhattacharya D. 2006. Defining the major lineages of red algae (Rhodophyta). *J Phycol.* 42(2):482–492.
- Yu G. 2020. Using ggtree to visualize data on tree-like structures. *Curr Protoc Bioinformatics* 69(1):e96.
- Zhang X, et al. 2019. The mitochondrial and chloroplast genomes of the green alga *Haematococcus* are made up of nearly identical repetitive sequences. *Curr Biol.* 29(15):R736–R737.

**Associate editor:** Dr. Daniel Sloan