Author/s:
Preuss, M;Verbruggen, H;West, JA;Zuccarello, GC

Title:
Divergence times and plastid phylogenomics within the intron-rich order Erythropeltales (Compsopogonophyceae, Rhodophyta)

Date:
2021-06-01

Citation:
Preuss, M., Verbruggen, H., West, J. A. & Zuccarello, G. C. (2021). Divergence times and plastid phylogenomics within the intron-rich order Erythropeltales (Compsopogonophyceae, Rhodophyta). Journal of Phycology, 57 (3), pp.1035-1044. https://doi.org/10.1111/jpy.13159.

Persistent Link:
https://hdl.handle.net/11343/298448

DR. HEROEN VERBRUGGEN (Orcid ID : 0000-0002-6305-4749)

DR. GIUSEPPE C. ZUCCARELLO (Orcid ID : 0000-0003-0028-7227)

Article type : Regular Article

Divergence times and plastid phylogenomics within the intron-rich order erythropeltales (Compsopogonophyceae, Rhodophyta)[1]

Maren Preuss
School of Biological Sciences, Victoria University of Wellington, PO Box 600, Wellington, 6140, New Zealand

Heroen Verbruggen
School of BioSciences, University of Melbourne, Parkville, VIC 3010, Australia

John A. West
School of BioSciences, University of Melbourne, Parkville, VIC 3010, Australia
and
Giuseppe C. Zuccarello [2]
School of Biological Sciences, Victoria University of Wellington, PO Box 600, Wellington, 6140, New Zealand

Running title: Phylogenomics Erythropeltales

[1] Received _____ Accepted _____

[2] Corresponding author: joe.zuccarello@vuw.ac.nz

Editorial Responsibility: O. De Clerck (Associate Editor)

ABSTRACT

The advent of high-throughput-sequencing (HTS) has allowed for the use of large numbers of coding regions to produce robust phylogenies. These phylogenies have been used to highlight relationships at ancient diversifications (subphyla, class), and highlight the evolution of plastid genome structure. The Erythropeltales are an order in the Compsopogonophyceae, a group with unusual plastid genomes but with low taxon sampling. We use HTS to produce near complete plastid genomes of all genera, and multiple species within some genera, to produce robust phylogenies to investigate character evolution, dating of divergence in the group, and plastid organization, including intron patterns. Our results produce a fully supported phylogeny of the genera in the Erythropeltales, and suggest that morphologies (upright versus crustose) have evolved multiple times. Our dated phylogeny also indicates that the order is very old (~ 800 Ma), with diversification occurring after the ice ages of the Cryogenian period (750-635 Ma). Plastid gene order is congruent with phylogenetic relationships and suggests that genome architecture does not change often. Our data also highlight the abundance of introns in the plastid genomes of this order. We also produce a nearly complete plastid genome of *Tsunamia transpacifica* (Stylonematophyceae) to add to the taxon sampling of genomes of this class. The use of plastid genomes clearly produces robust phylogenetic relationships that can be used to infer evolutionary events, and increased taxon sampling, especially in less well-known red algal groups, will provide additional insights into their evolution.

Key index words: ancestral state construction; chloroplast; comparative analysis of introns; deep phylogeny; diversification; group II introns; molecular clock; phylogenetics; plastid genome; red algae

Abbreviations: ARD, all rates different model; CTAB, cetyl trimethylammonium bromide; ER, equal rates model; HTS, High-throughput-sequencing; LCB, locally collinear block; SYM, symmetrical model

INTRODUCTION

The use of complete organelle genome data to accurately determine the relationships of algal groups has become more common with the advent of cheaper high-throughput-sequencing (HTS). These methods have been applied at several levels of taxonomic inquiry in red algae

from ordinal, family to generic relationships (e.g., Costa et al. 2016, Díaz-Tapia et al. 2017, 2019, Iha et al. 2018). These methods have been applied to non-florideophyceaen red algal orders (Qiu et al. 2016, Muñoz-Gómez et al. 2017) but taxon sampling is still low in some orders.

The red algal order Erythropeltales (also known as the Erythropeltidales but an incorrect spelling as the order needs to be typified by the generic name *Erythropeltis*; Athanasiadis 2016) is a commonly encountered group of mostly microscopic multicellular algae found worldwide. They include both simple filamentous, or occasionally monostromatic blades, and crustose growth forms (Zuccarello et al. 2010, 2011). Due to their simple morphology their taxonomy has been controversial, until the advent of molecular studies that have clarified some relationships, confirmed several genera (Zuccarello et al. 2010) and species (Zuccarello et al. 2011) and led to the description of new genera and species (e.g., West et al. 2012).

The Erythropeltales is in the class Compsopogonophyceae, one of the classes of the subphylum Proteorhodophytina, the others being the Porphyridiophyceae, the Rhodellophyceae, and the Stylonematophyceae (Muñoz-Gómez et al. 2017). Within the Compsopogonophyceae, there are two other orders; the Compsopogonales (with the common genus *Compsopogon*, among others) and the Rhodochaetales (containing the rare monotypic genus *Rhodochaete*). *Rhodochaete* was once thought to be a pivotal red alga, as it showed characters typical for florideophyceaen (pit plugs in all life stages) and features considered more primitive, such as simple sexual reproductive structures (Magne 1960, Pueschel and Magne 1987). Later studies of *Rhodochaete pulchella* (as *R. parvula*) from Australia (isolate JAW3924) suggested it was sister to the Erythropeltales (Zuccarello et al. 2000), excluding the Compsopogonales. The Compsopogonales is at present composed of two families (the Boldiaceae and the Compsopogonaceae), while the Erythropeltales contain the single family Erythrotrichiaceae.

The plastid genomes of the Proteorhodophytina have some unusual features (Muñoz-Gómez et al. 2017), including large numbers of introns: *Boldia erythrosiphon* with 20% of its 226 kb plastid genome consisting of introns; and 16% of the 221 kb genome of *Rhodochaete pulchella* (as *R. parvula*) consisting of introns, which are minimal compared to the amount of introns in some of the larger plastid genomes in the Proteorhodophytina (e.g., 63% of the 1.217 Mb genome of *Corynoplastis japonica*; Muñoz-Gómez et al. 2017). While gene content in these genomes was very similar, and within the bounds of other red algal plastid genomes, there were a few structural differences. For example, the plastid of *Boldia erythrosiphon*

(Compsopogonales) lacked a ribosomal RNA inverted repeat, while the plastid of *Rhodochaete pulchella* did have an inverted repeat, a common feature in the plastids under study (Muñoz-Gómez et al. 2017). Inverted repeats are found in many red algal plastids of the Florideophyceae, but it is not a ubiquitous feature (Lee et al. 2016), and differences are due to different inversion events leading to multiple plastid architecture-types throughout the class (Lee et al. 2016).

The aim of this study was to characterize the plastid genomes for the majority of described genera within the Erythropeltales and use them to: 1) reconstruct a robust phylogeny and infer growth forms evolution on the group, 2) estimate divergence times of genera and species, and 3) determine genome arrangements including characterization of introns if present within the Erythropeltales.

MATERIALS AND METHODS

DNA was extracted using the CTAB method from cultured isolates of members of the Erythropeltales: *Erythrocladia irregularis* (JAW culture number #4467), *Erythrotrichia carnea* (JAW4057), *E. foliiformis* (JAW4058), *E. longistipitata* (JAW4251), *E. welwitschii* (JAW4704), *Madagascaria erythrocladioides* (JAW4480), *Porphyropsis coccinea* (JAW4073), *Porphyrostromium boryanum* (JAW4706), *P. japonicum* (JAW4679), *Pseudoerythrocladia kornmannii* (JAW4464), *Sahlingia subintegra* (JAW4466) and one sample from the Stylonematophyceae: *Tsunamia transpacifica* (Table S1, Fig. S1 in the Supporting Information).

Extracted DNA of red algae was pooled with green algae DNA to reduce sequencing costs with a distinguishable alternate project organism (Díaz-Tapia et al. 2017, 2019) and libraries were sequenced on the Illumina HiSeq platform by Novogene, with 150 bp paired end reads. Plastid genomes were assembled and annotated as previously described (Preuss et al. 2020). Only partial plastid genomes were assembled as the number of rRNA copies and the possibility of inverted repeats of these rRNA copies made confirming circularity and completeness of the genomes difficult.

Protein coding genes were extracted, aligned and translated using the bacterial/plastid genetic code. All alignments were trimmed using the automated trim option in TrimAL (http://trimal.cgenomics.org). Additional taxa were selected by following the previous study of the Proteorhodophytina (Muñoz-Gómez et al. 2017). Alignments were concatenated by genes for phylogenetic analyses and either individual models for every gene were selected by using ModelFinder (Table S2 in the Supporting Information) or the C40 and C60 protein

mixture models were selected in IQ-Tree (Lanfear et al. 2012, Trifinopoulos et al. 2016). ML topology robustness was performed using 1000 UFbootstrap (Hoang et al. 2018) and SH-aLRT (Guindon et al. 2010) in IQ-tree. The Cyanidiophyceae were used as outgroup to root the tree.

Molecular clock analysis was inferred using Beast2 (Bouckaert et al. 2019) and the phylogeny produced by the C60 models was chosen as a fixed topology, as it was the only one confirming Eurhodophytina as sister to Proteorhodophytina. Each gene had an individually chosen site model with a gamma distribution category for four rate categories based on ModelTest in IQtree. The cpREV model was selected for all other genes where the best selected model is not present in Beast2. The molecular clock model and tree were linked across genes. For the analysis, the relaxed clock log model was selected and normal distribution of the *Bangiomorpha pubescens* Butterfield fossil at 1,198 ± 12 Ma (although this data could be a slight overestimates; Gibson et al. 2018) was used as a constraint for all red algae, excluding the Cyanidiales (Yang et al. 2016). The calibrated Yule model was selected for the tree prior. The topology was fixed by setting the weight to 0 for all operators and removing the "narrow-exchange", "subtree-slide", "wide-exchange" and "Wilson-balding" operators. The MCMC analysis were run for 1.5 million generations with sampling every 1000 generations. TreeAnnotator in Beast2 was used to create one consensus tree from all generated trees with a burn-in of 25%.

The dated phylogeny was used to reconstruct ancestral states of growth form using R version 4.0.2 software (R Core Team 2020) and the *geiger* and *phytools* package (Revell 2012, Pennell et al. 2014). All taxa within the Cyanidiales, Bangiales and Florideophyceae were removed and the remaining species were characterised as upright, unicellular or crustose. The ER (equal rates), SYM (symmetrical) and ARD (all rates different) models of multistate discrete character evolution were compared (models discussed in https://www.rdocumentation.org/packages/geiger/versions/2.0.7/topics/fitDiscrete). The ER model was the best model fit based on AIC and AICc parameters whereas the ARD model was the model best fitted based on the biological processes of growth form evolution. Trait mapping was performed 100 times and traits were combined on the fixed topology.

Progressive Mauve alignment was used to compare genome fragments across the Compsopogonophyceae. Specifically we used (1) the contigs ranging from *psb*V to *psb*D, (2) the contigs between *ycf*60 and *ycf*33, (3) the contigs ranging from *ycf*3 to *rpl*20 (Darling et al. 2004). *Erythrocladia irregularis* was removed from the Mauve alignment of contig 3 (*ycf*3-*rpl*20) because of its short length (11,101 bp). We also performed a separate alignment

between the complete genome of *Bangiopsis subsimplex* (KY709207) and the partial genome of *Tsunamia transpacifica.* Mauve alignments could not be compared between Compsopogonophyceae and Stylonematophyceae as plastid genomes were only partial and there were major rearrangements between these classes.

We carried out comparative analyses of introns across the Compsopogonophyceae plastid genomes, but excluded *Compsopogon caeruleus* because no annotated introns were present. Intron groups were determined for all contigs using genetic code 11 (bacterial, archael and plant plastid) in RNAweasel (https://megasun.bch.umontreal.ca/cgi-bin/RNAweasel/RNAweaselInterface.pl).

RESULTS

Phylogeny

The concatenated plastid data set, from our almost complete plastid genomes and published genomes, contained 39,051 amino acids of 171 protein coding genes from 29 taxa. ML topologies showed almost full support (100%; Fig. 1, Figs. S1-S3 in the Supporting Information) on all branches in the topology, only some of the deeper-branching relationships between classes are less than fully supported. One of the unsupported deep-branching relationships is the monophyly of the Proteorhodophytina, which was only recovered in the ML topology based on C60 models (but only with 39% support) but not in the ML trees where models were selected by ModelFinder and when using C40 models (Figs. S1-S2). The monophyly of the Compsopogonophyceae is fully supported in all analyses, as is the distinction between the Erythropeltales (including *Rhodochaete pulchella*) and the Compsopogonales (Fig. 1). The Erythropeltales are non-monophyletic as the order contains *Rhodochaete pulchella*, which is classified in the order Rhodochaetales. In the Erythropeltales, the upright genera with multiple species sampled (*Erythrotrichia* and *Porphyrostromium*) are monophyletic and, in this case, are sister genera to the crustose *Sahlingia subintegra*. There is also a fully supported grouping of the genera *Porphyropsis*, *Pseudoerythrocladia* and *Erythrocladia* (one upright and two crustose growth forms). Finally, *Madagascaria erythrocladioides* and *Rhodochaete pulchella* (a crust and an upright taxon) are a clade sister to all the other genera of Erythropeltales. *Boldia erythrosiphon* and *Compsopogon caeruleus* also form a supported clade (Fig. 1).

The estimated divergence times for Compsopogonophyceae, was during the late Proterozoic, around 844 (905-862) Ma. The individual classes, Porphyridiophyceae, Rhodellophyceae and the Stylonematophyceae, diverged in the Mesozoic, approximately 139

(154-130) Ma, 114 (122-107) Ma and 149 (159-140) Ma, respectively. Within the Compsopogonophyceae, the order Erythropeltales (including *Rhodochaete pulchella*) diverged in the late Proterozoic approximately 814 (831-781) Ma and the Compsopogonales around 628 (687-579) Ma. The Erythropeltales diversified during the late Proterozoic to Paleozoic after the Sturtian and Marinoan ice ages. The youngest genera/species in the Erythropeltales are estimated to have diverged during the Paleozoic and Mesozoic (Fig. 1).

The ancestral character state reconstruction showed some uncertainty about growth form near the deep-branching relationships of Stylonematophyceae, Rhodellophyceae, Compsopogonophyceae and Porphyridiophyceae with unicellularity slightly more likely using the all rates different (ARD) model. Filamentous upright is the most common estimated ancestral state along most of the internal nodes within the Erythropeltales. The crustose growth form evolved in three separate clades possibly during ice ages in the late Proterozoic in *Madagascaria* and *Sahlingia*, and the Mesozoic in *Erythrocladia-Pseudoerythrocladia* (Fig. 1).

Plastid genomes.

Partial plastid genomes were assembled in one contig (*Tsunamia transpacifica*), two contigs (*Erythrotrichia carnea, E. foliiformis, E. longistipitata, E. welwitschii, Madagascaria erythrocladioides, Porphyrostromium boryanum, P. japonicum, Pseudoerythrocladia kornmannii, Sahlingia subintegra*) or three contigs (*Erythrocladia irregularis*, *Porphyropsis coccinea*). The total assembled size ranged from 196,222 to 206,940 bp (between 93-98% of the complete plastid genome of *Erythrotrichia carnea* KX284721) containing 191-198 protein coding genes, 25-28 tRNAs, 0-3 rRNAs and 44-94 introns (Table 1). All taxa contained at least one intron, many genes had one intron (19-29 genes) or two introns (5-17 genes) depending on species, and occasionally multiple introns (3-9) per gene (Table S3). *Madagascaria erythrocladiodes* was the only taxon containing an intron with an opposite transcription direction from the gene it is contained in (*psa*F). Almost half of all taxa share the same intron position for a gene: 50% of all species = 9 genes, 75% = 14 genes, 100%= 3 genes (Table S3 in the Supporting Information).

In the Erythropeltales, 29-36 genes contain a total of 45-56 introns whereas both the number of genes containing introns and number of introns are higher in *Boldia erythrosiphon* and *Rhodochaete pulchella* (Table 2). Several genes with introns are only present in one taxon, e.g., *B. erythrosiphon* (*acp*P, *atp*B, *cb*6, *psb*B, *rpl*23, *sec*G, *suf*B, *orf*64, *psa*C), *R. pulchella* (*pet*B, *ycf*33, ORF62) and *E. irregularis* (ORF565; Tables S3-S4 in the Supporting

Information). Intron group-II-derived is the most common intron group type, but slightly over half of all introns are undefined (Tables 2, S4). Several annotated introns contained two intron matches within the same intron region (e.g., within the *atp*D gene of 8 species), indicating even though one intron is annotated, there may be two introns present (Tables 2, S4). Introns within *dna*K and *psa*A contained one or two protein-coding genes across all taxa (Tables 2, S4).

The Mauve alignment for the *psb*V to *psb*D contig (13,113-16,830 bp) identified four Locally Collinear Blocks (LCBs). Three LCBs were present in all taxa and one LCB (blue), in an intergenic space, was only present in *Pseudoerythrocladia kornmannii* and *Porphyrostromium japonicum* (Fig. S4 in the Supporting Information). Genes are not rearranged in this contig and the number of LCBs are likely be caused by big gaps of intergenic regions between LCBs and the sequence similarity of a small piece (blue LCB).

The Mauve alignment for the *ycf*60 to *ycf*33 contig (161,011-174,050 bp; 121,144 bp for *Porphyropsis coccinea*,) identified 11 LCBs (Fig. S5 in the Supporting Information). Two of the LCBs (purple, light green) in *Boldia erythrosiphon* and *Compsopogon caeruleus* are inverted to all other taxa. (Fig. S5A). Five LCBs are found within *dna*K in most taxa of the Erythropeltales with all of these containing a 'green' LCB and one of two 'red' LCBs. One 'purple' LCB is only present in *E. longistipitata* and *E. welwitschii*, and one 'yellow' LCB is only present in *E. carnea* (JAW4057) and *E. welwitschii*. All these LCBs within *dna*K do not seem to correlate with intron numbers or intron position but might be caused by the coded genes within the introns (Fig. 5B).

The Mauve alignment for the *ycf*3 to *rpl*20 contig (54,401-89,029 bp) identified 17 LCBs (Figs. S6 and S7 in the Supporting Information). All LCBs in *Boldia erythrosiphon* and *Compsopogon caeruleus* are transcribed in the same direction, without rearrangements (Figs. S6 and S7). All other included taxa show inversions and rearrangements compared to *Boldia-Compsopogon*. One LCB including part of *gro*EL and *syf*B (see LCB 8 in Fig. S6) can only be found in *Boldia erythrosiphon*, *Compsopogon caeruleus*, *Rhodochaete pulchella*, and *Madagascaria erythrocladioides*, which is inverted in the last two taxa compared to the first two. Another LCB also within *gro*EL (see LCB 15 in Fig. S6) was only observed in *Madagascaria erythrocladioides*, *Porphyropsis coccinea*, *Pseudoerythrocladia kornmannii* and *Rhodochaete pulchella* (Figs. S6 and S7).

Stylonematophyceae.

The partial plastid genome of *Tsunamia transpacifica* was assembled in one contig of 206,801 bp containing 196 protein coding genes, 15 introns, 3 rRNA and 30 tRNA's (Table 1) but we were not able to confirm circularity. The plastid genome of *T. transpacifica* is similar in genome size, gene content and gene arrangement, but contains an extra tRNA intron than the complete plastid genome of *Bangiopsis subsimplex* (205,002 bp long, KY709207). Although, overall it has less introns (15 vs. 25) than *B. subsimplex* (Table 1, Fig. S8 in the Supporting Information).

DISCUSSION

Our nearly complete plastid genomes of several genera and species of the Erythropeltales and our phylogenetic reconstructions clearly show the utility of plastid phylogenomics in producing robust and useful evolutionary relationships. These phylogenies are very similar to the relationships seen from previous studies (Zuccarello et al. 2010, 2011) using a few genes, but with full support and therefore more confidence in evolutionary patterns.

Plastid genome structure also closely matches phylogenetic relationships, with the groupings: *Madagascaria erythrocladiodes/Rhodochaete pulchella*; *Porphyropsis coccinea/Pseudoerythrocladia kornmannii*; and the remaining Erythropeltales (*Porphyrostromium, Sahlingia, Erythrotrichia*) having near identical genome structure. Plastid genome rearrangements are rare in this group.

The estimated divergence time of the Compsopogonophyceae, based on sampling of the majority of the genera, was during the late Proterozoic, around 884 (905-862) Ma, with continued diversification from the late Proterozoic until the Mesozoic. These early divergence estimates are older than the diversification of other classes (e.g., Bangiophyceae- 468 Ma; Porphyridiophyceae- 139 Ma). Early studies using smaller data sets for the Compsopogonophyceae, calculated their divergence times at approximately 300 Ma using only 7 genes, 547 Ma for whole plastid genomes but also 1195 Ma for whole mitochondrial genomes but with fewer samples (Yang et al. 2016, Nan et al. 2017). This apparent ancient divergence during the late Proterozoic (~ 884 Ma) is on the order of divergence times for the Florideophyceae (~ 860 Ma) and yet the morphological complexity of these two multicellular groups is markedly different, with high species numbers and morphological complexity in the Florideophyceae, in comparison to the current group under study.

These time diversification estimates are very dependent on the fossil calibrations used. The affiliation of Proterozoic fossils (> ~500 Ma) to extant groups is problematic, due their scarcity and generalized body plans. Crown group rhodophytes have been proposed for fossils

from ~1.6 billion year ago (Zhang 1989, Bengtson et al. 2017). These fossils have very generalized growth forms (filaments, or pseudo-parenchyma so called 'cell fountains') but lack other characters (pit connections). Using macrofossil stem-group designations (e.g., Florideophyceae) for these 1.6 billion year old fossils makes dates of the origins of the rhodophytes, including the Compsopogonophyceae, improbably old (Berney and Pawlowski 2006, Parfrey et al. 2011). The use of the *Bangiomorpha* fossil as a stem group Rhodophyta (excluding the Cyanidiales) is conservative, and has been used previously (see above), but the true affiliations of these Proterozoic fossils remain uncertain.

While comparative genomic studies have not been conducted in the Compsopogonophyceae, aside from the present study, genomic studies comparing Florideophyceae with Porphyridiophyceae (one species) show that transposable elements are greatly increased in Florideophyceae (Lee et al. 2018). Transposable element activity is known to contribute to gene modification, rearrangement and regulation in plants which may influence morphological variability (Bennetzen and Wang 2014, Wendel et al. 2016). Of course, sexual reproduction also affects transposable element number and activity (Arkhipova 2005), an area of research that is still lacking in the Compsopogonophyceae, due to incomplete life cycle observations (summarized in Zuccarello et al. 2010, 2011), rarity of sexual reproduction or even its loss. How genomic changes may have shaped the morphological evolution of this group still needs to be addressed.

Estimated divergence times are also very old for the Erythropeltales, with the order originating approximately 814 Ma, and diversifying 724-573 Ma just after the great Cryogenian glaciations (720-635 Ma; Fairchild and Kennedy 2007). The ancestors of the order obviously survived these world-wide climate changes, and the warming of the earth after these glaciations allowed for diversification of the different lineages (*Sahlingia*, *Erythrotrichia*, *Porphyriostromium*). Diversification and evolution of different morphological growth forms after the Cryogenian glaciations in green algae was associated with novel grazing and habit pressure to outcompete other organisms by growing upright (Cortona et al. 2020). Upright habit is the most likely growth form of the ancestral Erythropeltales, based on our reconstruction. Whether these late Proterozoic earth warmings lead to diversification in other red algal groups needs a wider analysis but may show the importance of these eras in algal diversification and possible historical biogeography.

Interestingly, the genus *Erythrotrichia*, with a simple filamentous morphology, has persisted for over 500 Ma. This status is maybe not remarkable as the most ancient fossil recognizable to a present day 'genus' is the red alga *Bangiomorpha pubescens* that closely

resembles present day *Bangia* (Butterfield 2000). While *Bangiomorpha pubescens* was originally placed as an ancestor of the Bangiales, even Butterfield (2000) noticed similarity to other red algal groups, including the Erythropeltales, more conservatively placing its position, as in the present study, at the base of non-Cyanidiophycean red algae. The *Bangiomorpha* illustrations of holdfasts (Fig. 6; Butterfield 2000) do resemble the multicellular bases of some Erythropeltales (Zuccarello et al. 2011). An intriguing insight is the identification of potential multicellular sexual structures (carpospores) in *Bangiomorpha* which are not seen in Erythropeltales, which either have simple unicellular reproductive structure (in *Rhodochaete pulchella* for example, Magne 1960) or lack any known alternation in life cycles. This suggests that sexuality, and sexual structures, could have been lost or simplified in this order from their potential common ancestor.

This potential simplification from a more complex ancestor is even seen in vegetative morphology, as some Proteorhodophytina have even simplified to unicellular vegetative cells, although not in the Erythropeltales. Ancestral character reconstruction based on our phylogenetic results showed that the upright and crustose forms have developed multiple times, at least three, within the Erythropeltales. The link between these two forms is unsurprising as many genera produce disk-like attachment structures early in development from which uprights develop. For example, *Porphyropsis coccinea* produces a crust very similar to the crustose *Erythrocladia* sp. from which uprights then develop (Murray et al. 1972, Kornmann and Sahling 1985). As mentioned previously, *Porphyrostromium* produces a crust which morphologically is very similar to *Sahlingia* from which the upright develops (Kornmann and Sahling 1985, Kikuchi and Shin 2011). Interestingly, our phylogeny supports this, as *Porphyropsis* is closely related to *Erythrocladia*, while *Porphyrostromium* shares a more recent common ancestor to *Sahlingia* than to the other two genera mentioned. This developmental morphology is further complicated by the supposition that in some taxa (e.g., *Porphyrostromium japonicum*) the two morphologies could be alternate phases of the same life history, with a solely crust phase being the diploid and the crust/upright the gametophyte (Kikuchi and Shin 2011). While simple sexual structures have been reported, though rarely, in the Erythropeltales (e.g., West et al. 2012), complete life cycles especially evidence of meiosis, or chromosome counts, is still lacking and confirmation of sexual life history stages is clearly needed.

Intron-richness is one of the unique features within the Compsopogonophyceae plastids. The high number of introns are predominantly classified as intron II (derived). All our compared taxa also contain coding genes (maturases) in their *dna*K and *psa*A introns and

these are mostly in the same position between the majority of compared taxa, indicating that these are ancient introns. Introns can be classified as group I or group II introns but currently only group II introns have been identified in plastid genomes of red algae (Perrineau et al. 2015, Muñoz-Gómez et al. 2017). One possible functional advantage of retaining these introns is the ability for alternative splicing of RNA, as intron group II can be spliced by maturases encoded in the introns (Wank et al. 1999). Another explanation as to why introns have been maintained over time may be explained by the Hill-Robertson inference, that states recombination between two sites will be increased by length of introns (Comeron and Kreitman 2000), which is strongest in asexual organisms with completely linked genomes (Whitlock et al. 2016). Variation of plastid genome architecture is relatively small in the Compsopogonophyceae indicating little to no recombination between these plastid genomes. Whatever the reasons for intron retention, the Compsopogonophyceae are an intriguing group to investigate intron functionality in depth and will contribute to our understanding of intron evolution in red algae.

Our results show several clear groupings within the Erythropeltales. The order as presently designated also contains *Rhodochaete pulchella*, if *Madagascaria erythrocladioides* is included, therefore making the Erythropeltales paraphyletic. *Rhodochaete pulchella* has an interesting history, as it was once suggested to be an important "intermediate" species between 'primitive' and 'advanced' red algae (Garbary and Gabrielson 1990). This was due the presence of apical growth and pit-connections in all life stages, considered advanced characters, and simple undifferentiated one-celled male and female gametangia, considered primitive characters (Magne 1960, Boillot 1975). The pivotal position was questioned once an isolate matching the description of *Rhodochaete* (branched uniseriate filament, non-stellate plastid,) was collected and cultured from Australia, as the type culture from France had been lost. The Australian isolate only reproduced asexually by monospores (Zuccarello et al. 2000). The intermediate (ancestral) status of any extant species is always a questionable hypothesis, and morphological characters that are considered diagnostic and unique are found, using independent phylogenetic reconstruction methods (i.e., DNA sequence data), to have often evolved multiple times in lineages. For example, in the Laminariales (meristem splitting) (Lane et al. 2006), or green algal order Volvocales (isogamy; Nozaki et al. 2000) have evolved multiple times. Molecular evidence on the Australian isolate of *Rhodochaete pulchella* includes both the original sequences (ribosomal RNA), other nuclear genes, plus complete plastid sequences. While it is tempting to designate this culture the type, and therefore synonymize the order Rhodochaetales into the Erythropeltales, we must bear in

mind that the original culture was from Mediterranean France. The Australian isolate does not show, in culture, sexual reproductive structures, plus the characters identifying these algae of simple morphologies, in this case a branched filament, could be homoplasious; and several other characteristics also do not conform to the original description (e.g., cell dimensions; Mateo-Cid et al. 2002). We suggest an intensive collection effort in the Mediterranean be undertaken to clearly show that the samples from the type area are the same as the Australian sample.

We have produced a fully supported phylogeny of the Compsopogonophyceae and especially the Erythropeltales using plastid genomic data. Our data sampled all the genera, and several species within genera, and shows that the group has a long evolutionary history with fairly stable plastid genome structure, especially within the Erythrotrichiaceae. The position and number of introns also is conserved, and the forces maintaining this extra genome burden need investigation. Our study also suggests that the order Rhodochaetales needs further careful investigation to confirm or dismiss its previous transitional position in red algal evolution.

ACKNOWLEDGEMENTS

This research was partial supported by the Marsden Fund Council from Government funding, administered by the Royal Society of New Zealand. This work was supported *via* strategic research funds from the School of Biological Sciences at Victoria University of Wellington.

REFERENCES

Arkhipova, I.R. 2005. Mobile genetic elements and sexual reproduction. *Cytogenet. Genome Res.* 110:372–82.

Athanasiadis, A. 2016. Phycologia Europaea Rhodophyta Vol. I. Thessaloniki. 1–762 pp.

Bengtson, S., Sallstedt, T., Belivanova, V. & Whitehouse, M. 2017. Three-dimensional preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae. *PLoS Biol.* 15:e2000735.

Bennetzen, J.L. & Wang, H. 2014. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* 65:505–30.

Berney, C. & Pawlowski, J. 2006. A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc R Soc Biol Ser B*. 273:1867–72.

Boillot, A. 1975. Cycle biologique de *Rhodochaete parvula* (Thuret) (Rhodophycées, Bangiophycidées). *Publicazione Zoolologia Stn. Napoli*. 39:67–83.

437 Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina,
438 A., Heled, J., Jones, G., Kühnert D., De Maio, N., Matschiner, M., Mendes, F.K.,
439 Müller, N.F., Ogilvie, H.A., de Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D.,
440 Siveroni, I., Suchard, M.A., Wu, C. H., Xie, D, Zhang, C., Stadler, T. & Drummond,
441 A.J. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary
442 analysis. *PLoS Comput. Biol*. 15:e1006650.

443 Butterfield, N.J. 2000. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution
444 of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes.
445 *Paleobiol*. 26:386–404.

446 Comeron, J.M. & Kreitman, M. 2000. The correlation between intron length and
447 recombination in *Drosophila*. Dynamic equilibrium between mutational and selective
448 forces. *Genetics* 156:1175–90.

449 Costa, J.F., Lin, S. M., MacAya, E.C., Fernández-Garciá, C. & Verbruggen, H. 2016.
450 Chloroplast genomes as a tool to resolve red algal phylogenies: a case study in the
451 Nemaliales. *BMC Evol. Biol.* 16:1–13.

452 Darling, A.C.E., Mau, B., Blattner, F.R. & Perna, N.T. 2004. Mauve: Multiple alignment of
453 conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–403.

454 Del Cortona, A. , Jackson, C.J., Bucchini, F., Van Bel, M., D'hondt, S., Skaloud, P.,
455 Delwiche, C.F., Knoll, A.H., Raven, J.A., Verbruggen, H. Vandepoele, K., De Clerck, O.
456 & Leliaert, F. 2020. Neoproterozoic origin and multiple transitions to macroscopic
457 growth in green seaweeds. *Proc. Natl. Acad. Sci. USA* 117:2551–9.

458 Díaz-Tapia, P., Maggs, C.A., West, J.A. & Verbruggen, H. 2017. Analysis of chloroplast
459 genomes and a supermatrix inform reclassification of the Rhodomelaceae (Rhodophyta).
460 *J. Phycol.* 53:920–37.

461 Díaz-Tapia, P., Pasella, M.M., Verbruggen, H. & Maggs, C.A. 2019. Morphological evolution
462 and classification of the red algal order Ceramiales inferred using plastid phylogenomics.
463 *Mol. Phylogenet. Evol.* 137:76–85.

464 Fairchild, I.J. & Kennedy, M.J. 2007. Neoproterozoic glaciation in the Earth system. *J. Geol.*
465 *Soc. London.* 164:895–921.

466 Garbary, D.J. & Gabrielson, P.W. 1990. Taxonomy and evolution. *In* Cole, K. M. & Sheath,
467 R. G. [Eds.] *Biology of the Red Algae*. Cambridge University Press, New York, pp. 477–
468 98.

469 Gibson, T.M., Shih, P.M., Cumming, V.M., Fischer, W.W., Crockford, P.W., Hodgskiss,
470 M.S.W., Wörndle, S., Creaser, R.A., Rainbird, R.H., Skulski, T.M. & Halverson, G.P.

2018. Precise age of *Bangiomorpha pubescens* dates the origin of eukaryotic photosynthesis. *Geology* 46:135–8.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* 59:307–21.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. 2018. UFBoot2: improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35:518–22.

Iha, C., Grassa, C.J., Lyra, G. de M., Davis, C.C., Verbruggen, H. & Oliveira, M.C. 2018. Organellar genomics: a useful tool to study evolutionary relationships and molecular evolution in Gracilariaceae (Rhodophyta). *J. Phycol.* 54:775–87.

Kikuchi, N. & Shin, J. A. 2011. *Porphyrostromium japonicum* (Tokida) Kikuchi comb. nov. (Erythropeltidales, Rhodophyta) from Japan. *Phycologia* 50:122–31.

Kornmann, P. & Sahling, P.H. 1985. Erythropeltidaceen (Bangiophyceae, Rhodophyta) von Helgoland. *Helgoländer Meeresuntersuchungen*. 39:213–36.

Lane, C.E., Mayes, C., Druehl, L.D. & Saunders, G.W. 2006. A multi-gene molecular investigation of the kelp (Laminariales, Phaeophyceae) supports substantial taxonomic re-organization. *J. Phycol.* 42:493–512.

Lanfear, R., Calcott, B., Ho, S.Y.W. & Guindon, S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* 29:1695–701.

Lee, J., Cho, C.H., Park, S.I., Choi, J.W., Song, H.S., West, J.A., Bhattacharya, D. & Yoon, S.H. 2016. Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. *BMC Biol.* 14:75.

Lee, J.M., Yang, E.C., Graf, L., Yang, J.H., Qiu, H., Zelzion, U., Chan, C.X., Stephens, T.G, Weber, A.P.M., Boo, G.H., Boo, S.M., Kim, K.M., Shin, Y., Jung, M., Lee. S.J., Yim. H. S., Lee. J. H, Bhattacharya, D., & Yoon, H.S. 2018. Analysis of the draft genome of the red seaweed *Gracilariopsis chorda* provides insights into genome size evolution in Rhodophyta. *Mol. Biol. Evol.* 35:1869–86.

Magne, F. 1960. Le *Rhodochaete parvula* Thuret (Bangioidée) et sa reproduction sexuée. *Cah. Biol. Mar.* 1:407–20.

Mateo-Cid, L.E., Mendoza-González, A.C. & Searles, R.B. 2002. New mexican records of marine algae including *Crouania mayae* sp. nov. (Ceramiaceae, Rhodophyta). *Caribb. J. Sci.* 38:205–21.

Muñoz-Gómez, S.A., Mejía-Franco, F.G., Durnin, K., Colp, M., Grisdale, C.J., Archibald,

505 J.M. & Slamovits, C.H. 2017. The new red algal subphylum Proteorhodophytina
506 comprises the largest and most divergent plastid genomes known. *Curr. Biol.* 27:1677-
507 1684.e4.
508 Murray, S.N., Dixon, P.S. & Scorrf, J.L. 1972. The life history of *Porphyropsis coccinea* var.
509 *dawsonii* in culture. *Br. Phycol. J.* 7:323–33.
510 Nan, F., Feng, J., Lv, J., Liu, Q., Fang, K., Gong, C. & Xie, S. 2017. Origin and evolutionary
511 history of freshwater Rhodophyta: Further insights based on phylogenomic evidence.
512 *Sci. Rep.* 7:2934.
513 Nozaki, H., Misawa, K., Kajita, T., Kato, M., Nohara, S. & Watanabe, M.M. 2000. Origin and
514 evolution of the colonial Volvocales (Chlorophyceae) as inferred from multiple,
515 chloroplast gene sequences. *Mol. Phylogenet. Evol.* 17:256–68.
516 Parfrey, L.W., Lahr, D.J.G., Knoll, A.H. & Katz, L.A. 2011. Estimating the timing of early
517 eukaryotic diversification with multigene molecular clocks. *Proc. Natl. Acad. Sci. USA*
518 108:13624–9.
519 Pennell, M.W., Eastman, J.M., Slater, G.J., Brown, J.W., Uyeda, J.C., Fitzjohn, R.G., Alfaro,
520 M.E. & Harmon, L.J. 2014. Geiger v2.0: An expanded suite of methods for fitting
521 macroevolutionary models to phylogenetic trees. *Bioinformatics*. 30:2216–8.
522 Perrineau, M.M., Price, D.C., Mohr, G. & Bhattacharya, D. 2015. Recent mobility of plastid
523 encoded group II introns and twintrons in five strains of the unicellular red alga
524 *Porphyridium*. *PeerJ*. 2015:e1017.
525 Preuss, M., Verbruggen, H. & Zuccarello, G.C. 2020. The organelle genomes in the
526 photosynthetic red algal parasite *Pterocladiophila hemisphaerica* (Florideophyceae,
527 Rhodophyta) have elevated substitution rates and extreme gene loss in the plastid
528 genome. *J. Phycol.* 56:1006–18.
529 Pueschel, C.M. & Magne, F. 1987. Pit plugs and other ultrastructural features of systematic
530 value in *Rhodochaete parvula* (Rhodophyta, Rhodochaetales). *Cryptogam. Algol.* 8:201–
531 9.
532 Qiu, H., Yoon, H.S. & Bhattacharya, D. 2016. Red algal phylogenomics provides a robust
533 framework for inferring evolution of key metabolic pathways. *PLoS Curr.* 8.
534 R Core Team 2020. R: A language and environment for statistical computing.
535 Revell, L.J. 2012. phytools: An R package for phylogenetic comparative biology (and other
536 things). *Methods Ecol. Evol.* 3:217–23.
537 Trifinopoulos, J., Nguyen, L.T., von Haeseler, A. & Minh, B.Q. 2016. W-IQ-TREE: a fast
538 online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44:W232–

5.
Wank, H., SanFilippo, J., Singh, R.N., Matsuura, M. & Lambowitz, A.M. 1999. A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. *Mol. Cell*. 4:239–50.
Wendel, J.F., Jackson, S.A., Meyers, B.C. & Wing, R.A. 2016. Evolution of plant genome architecture. *Genome Biol.* 17.
West, J.A., De Göer, S.L. & Zuccarello, G.C. 2012. Upright Erythropeltidales (Rhodophyta) in Brittany, France and description of a new species, *Erythrotrichia longistipitata*. *Cah. Biol. Mar.* 53:255–70.
Whitlock, A.O.B., Peck, K.M., Azevedo, R.B.R. & Burch, C.L. 2016. An evolving genetic architecture interacts with Hill–Robertson interference to determine the benefit of sex. *Genetics*. 203:923–36.
Yang, E.C., Boo, S.M., Bhattacharya, D., Saunders, G.W., Knoll, A.H., Fredericq, S., Graf, L. et al. 2016. Divergence time estimates and the evolution of major lineages in the florideophyte red algae. *Sci. Rep.* 6:21361.
Zhang, Y. 1989. Multicellular thallophytes with differentiated tissues from Late Proterozoic phosphate rocks of South China. *Lethaia*. 22:113–32.
Zuccarello, G., West, J., Bitans, A. & Kraft, G. 2000. Molecular phylogeny of *Rhodochaete parvula* (Bangiophycidae, Rhodophyta). *Phycologia* 39:75–81.
Zuccarello, G.C., Kikuchi, N. & West, J.A. 2010. Molecular phylogeny of the crustose Erythropeltidales (Compsopogonophyceae, Rhodophyta): new genera *Pseudoerythrocladia* and *Madagascaria* and the evolution of the upright habit. *J. Phycol.* 46:363–73.
Zuccarello, G.C., Yoon, H.S., Kim, H., Sun, L., de Goër, S.L. & West, J.A. 2011. Molecular phylogeny of the upright Erythropeltidales (Compsopogonophyceae, Rhodophyta): multiple cryptic lineages of *Erythrotrichia carnea*. *J. Phycol.* 47:627–37.

Fig. 1. Bayesian tree with estimated divergence times and estimated ancestral characters states, calculated using a relaxed log normal clock model in BEAST2 based on a fixed topology of ML phylogeny using C60 models and 171 plastid genes. Culture number of all isolates are given. Values of 100% bootstrap support of the ML phylogeny of C60 models are indicated as asterisk and values of <80% bootstrap are not shown. Node age is represented below the 95% confidence interval. The 95% confidence interval of each node bar is indicated as a grey bar and values are represented below within brackets. The placement of the fossil

constraint (i.e., *Bangiomorpha pubescens*) is indicated by an A. Branch lengths are proportioned to divergence times and lengths of different eons and eras are indicated below scale bar. Time frame of major ice ages are highlighted in grey: Sturtian (750-700 Ma), Marinoan (650-635 Ma), Andean-Saharan (460-430 Ma) and Karoo (360-260 Ma). Ancestral character state was calculated using ARD models in R and fixed topology. Characters states of growth form are upright (black), unicellular (white) and crustose (grey).

Table 1. Characterization of plastid genomes of the Compsopogonophyceae and Stylonematophyceae. Total contig size, number of assembled contigs, protein coding genes (as CDS), non-coding (nc)RNAs, rRNAs and tRNAs.

	Contig size (bp)	Contigs	CDS	ncRNA	rRNA	tRNA	Reference
Compsopogonophyceae							
Boldia erythrosiphon	226,658	1*	187	-	3	31	Muñoz-Gómez et al. 2017
Compsopogon caeruleus	221,013	1*	195	-	5	29	Nan et al. 2017
Erythrocladia irregularis	202,975	3	191	1	-	28	This study
Erythrotrichia carnea CCMP3225	210,691	1*	191	-	6	32	Lee et al. 2016
Erythrotrichia carnea JAW4057	196,222	2	195	1	-	28	This study
Erythrotrichia foliiformis	198,213	2	196	1	3	28	This study
Erythrotrichia longistipitata	200,029	2	196	1	-	28	This study
Erythrotrichia welwitschii	197,515	2	197	1	1	28	This study
Madagascaria erythrocladioides	206,940	2	195	1	-	28	This study
Porphyropsis coccinea	197,738	3	195	1	2	25	This study
Porphyrostromium boryanum	197,623	2	194	1	1	27	This study
Porphyrostromium japonicum	198,702	2	196	1	2	27	This study
Pseudoerythrocladia kornmannii	204,391	2	196	1	-	28	This study
Rhodochaete parvula	221,665	1*	195	-	6	31	Lee et al. 2016
Sahlingia subintegra	203,045	2	198	2	1	28	This study
Stylonematophyceae							

Bangiopsis subsimplex	205,002	1*	198	-	3	29	Muñoz-Gómez et al. 2017
Tsunamia transpacifica	206,801	1	196	1	3	30	This study

*complete circular plastid genome, [a]=introns are present in the genome but not annotated

Table 2. Total number of genes containing introns and total number of introns within the plastid genome of the Compsopogonophyceae. All introns were classified as either intron group II (derived), intron group II (domain V) or unidentified if no hit was found. Name of genes containing an intron were two RNA structures were identified is number of introns and containing genes.

	Total number of genes with introns	Total number of introns	Intron group II, derived	Intron group II (domain V)	Intron group: undefined	Genes containing one intron with two intron matches	Total number of introns containing genes
Erythrotrichia longistipitata	36	55	28	3 (*rpo*A: 2, *ycf*20: 1)	28	*atp*D, *rpo*A	3 (*dna*K: 2, *psa*A: 1)
Erythrotrichia carnea JAW4057	35	56	39		18	*bas*1	3 (*dna*K: 2, *psa*A: 1)
Erythrotrichia welwitschii	36	56	35		23	*atp*D, *atp*F	4 (*dna*K: 3, *psa*A: 1)

Erythrotrichia carnea CCMP3225	26	58	31	1 (*dna*K)	27	*atp*D	3 (*dna*K: 2, *psa*A: 1)
Erythrotrichia foliiformis	36	59	30	1 (*atp*I)	28	*atp*D, *bas*1	3 (*dna*K: 2, *psa*A: 1)
Porphyrostromium japonicum	34	56	24		32		3 (*dna*K: 2, *psa*A: 1)
Porphyrostromium boryanum	35	56	30	2 (*psa*F, *rpo*A	24	*atp*D	3 (*dna*K: 2, *psa*F: 1)
Sahlingia subintegra	32	52	36	1 (*atp*I)	16	*atp*D, *atp*F	3 (*dna*K: 2, *psa*A :1)
Porphyropsis coccinea	29	45	28		19	*atp*D, *bas*1	3 (*dna*K: 2, *psa*A: 1)
Pseudoerythrocladia kornmannii	34	55	37	2 (*psa*A, *psa*F)	18	*bas*1, *rps*7	3 (*dna*K: 2, *psa*A: 1)

Erythrocladia irregularis	33	55	32		26	*atp*D, *atp*F, *bas*1	2 (*dna*K: 1, *psa*A: 1)
Rhodochaete pulchella	41	64	45	3 (*dna*K, *pet*B, *ycf*33)	18	ORF62	5 (*dna*K: 2, *pet*B: 1, *psa*A: 1, *ycf*33: 1)
Madagascaria erythrocladioides	32	45	25		21	*atp*D	3 (*dna*K: 2, *psa*A: 1)
Boldia erythrosiphon	50	94	66	2 (*apc*D, *inf*C)	33	*atp*F, *cb*6, *dna*K, *rpo*A, *rps*2, *rps*7, *suf*C	2 (*dna*K: 1, *psa*A: 1)

Table S1. Accession numbers of plastid genomes with culture numbers, in alphabetical order by class.

Table S2. Models selected for individual plastid genes using ModelFinder in IQ-tree for phylogenetic analyse in IQ-Tree.

Table S3. Comparison of plastid genomes that contain introns between 14 taxa within the Compsopogonophyceae. Number indicates the number of introns, '+' indicates that the gene in present (annotated) but without introns, and '-' indicates the that the gene was not annotated within partial or complete plastid contigs. Intron positions are determined using ungapped distances of the *Boldia erythrosiphon* sequence as a reference. In absence of introns in genes of *B. erythrosiphon* an alternative taxon is used as a reference: [1] *Rhodochaete pulchella*, [2] *Porphyrostromium japonicum* and [3] *Erythrocladia irregularis*. Intron positions were not determined for *rpo*A due to problematic alignment. Extension of intron position indicates that the intron is inserted between 2 codons ('-0'), after the first codon ('-1') and after the third codon ('-2').

Table S4. Identified RNA structures were identified in introns using RNAweasel. V followed by e-values in brackets are classified as intron II (domain V) and only e-values in brackets are classified as intron II. Introns without a RNA structure hit are classified as undefined and introns that contain two identified RNA structures are bold. Name of coding gene within an intron region are given below the e-values.

Table S4. Identified RNA structures were identified in introns using RNAweasel. V followed by e-values in brackets are classified as intron II (domain V) and only e-values in brackets are classified as intron II. Introns without a RNA structure hit are classified as undefined and introns that contain two identified RNA structures are bold. Name of coding gene within an intron region are given below the e-values.

Fig. S1. ML topology of 171 concatenated plastid gene protein sequences where models were selected for every gene based on ModelFinder in IQtree (Table S2). Taxa represented species within the Bangiophyceae, Compsopogonophyceae, Florideophyceae, Rhodellophyceae and Stylonematophyceae (Table S1). Culture number of all isolates are given. *Cyanidium*

caldarium and *Galdieria sulphuraria* were used as outgroups. Values <80% SH-aLRT and <95% UFbootstrap are not shown. Asterisk represents full (100%) support.

Fig. S2. ML topology of 171 concatenated plastid gene protein sequences using C40 models. Taxa represented species within the Bangiophyceae, Compsopogonophyceae, Florideophyceae, Rhodellophyceae and Stylonematophyceae (Table S1). Culture number of all isolates are given. *Cyanidium caldarium* and *Galdieria sulphuraria* were used as outgroups. Values <80% bootstrap are not shown. Asterisk represents full (100%) support.

Fig. S3 ML topology of 171 concatenated plastid gene protein sequences using C60 models. Taxa represented species within the Bangiophyceae, Compsopogonophyceae, Florideophyceae, Rhodellophyceae and Stylonematophyceae (Table S1). Culture number of all isolates are given. *Cyanidium caldarium* and *Galdieria sulphuraria* were used as outgroups. Values <80% bootstrap are not shown. Asterisk represents full (100%) support.

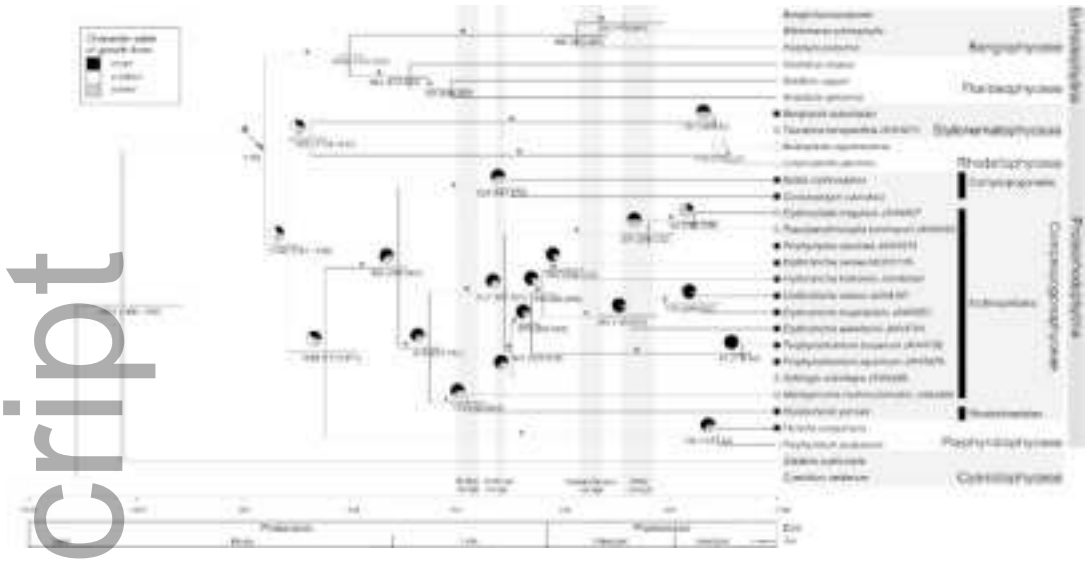
Fig. S4. Progressive Mauve alignment of the *psb*V to *psb*D contig between 15 taxa within the Compsopogonophyceae identified four LCB, with two independent inversions below the center line in a region without a gene annotation (blue LCB, arrow) in *Porphyrostromium japonicum* and *Pseudoerythrocladia kornmanii*. The three LCB present in all taxa read above the center line.

Fig. S5. Progressive Mauve alignment of the *ycf*60 to *ycf*33 contig between 15 taxa within the Compsopogonophyceae identified eleven LCB. A) Two LCB (purple and light green= indicated by arrows) in *Boldia erythrosiphon* and *Compsopogon caeruleus* (Comsopogonales) are inverted to all other taxa represented. B) Five LCB within the *dna*K are only present in most Erythropeltidales (black box)

Fig. S6. Schematic representation of the synteny of 17 identified LCBs within the *ycf*3 to *rpl*20 contigs between Compsopogonophyceae. Shared vertical grey block represents identical order of LCBs. Directions of arrow indicates LCB transcription direction (left arrow above center line or right arrow below center line).

Fig. S7. Progressive Mauve alignment of the *ycf*3 to *rpl*20 contig between 14 taxa within the Compsopogonophyceae identified 17 LCB. The light green LCB (arrow) can only be found in *Boldia erythrosiphon*, *Compsopogon caeruleus*, *Rhodochaete pulchella*, *Madagascaria erythrocladioides*. The purple LCB (arrow) was only observed in *Madagascaria erythrocladioides*, *Porphyropsis coccinea*, *Pseudoerythrocladia kornmannii*, *Rhodocaete pulchella*.

Fig. S8. Progressive Mauve alignment of the complete plastid genome of *Bangiopsis subsimplex* and the partial plastid genome of *Tsunamia transpacifica* shows one LCB block indicating the similarity between these two plastid genomes.



jpy_13159-20-248_f1.tif