Red Algal Mitochondrial Genomes Are More Complete than Previously Reported

Eric D. Salomaki* and Christopher E. Lane

Department of Biological Sciences, University of Rhode Island

*Corresponding author: E-mail eric.salomaki@gmail.com.

Accepted: November 7, 2016

Data deposition: Newly generated data has been deposited in GenBank. The mitochondrial genomes of *Gracilariopsis andersonii*, *Gracilariophila oryzoides*, *Choreocolax polysiphoniae* and *Vertebrata lanosa* are listed under accessions KX687878, KX687879, KX687877, KX687880 respectively, and the *Ahnfeltia plicata atp8* gene is under accession KX687876.

Abstract

The enslavement of an alpha-proteobacterial endosymbiont by the last common eukaryotic ancestor resulted in large-scale gene transfer of endosymbiont genes to the host nucleus as the endosymbiont transitioned into the mitochondrion. Mitochondrial genomes have experienced widespread gene loss and genome reduction within eukaryotes and DNA sequencing has revealed that most of these gene losses occurred early in eukaryotic lineage diversification. On a broad scale, more recent modifications to organelle genomes appear to be conserved and phylogenetically informative. The first red algal mitochondrial genome was sequenced more than 20 years ago, and an additional 29 Florideophyceae mitochondria have been added over the past decade. A total of 32 genes have been described to have been missing or considered non-functional pseudogenes from these Florideophyceae mitochondria. These losses have been attributed to endosymbiotic gene transfer or the evolution of a parasitic life strategy. Here we sequenced the mitochondrial genomes from the red algal parasite *Choreocolax polysiphoniae* and its host *Vertebrata lanosa* and found them to be complete and conserved in structure with other Florideophyceae mitochondria. This result led us to resequence the previously published parasite *Gracilariophila oryzoides* and its host *Gracilariopsis andersonii*, as well as reevaluate reported gene losses from published Florideophyceae mitochondria. Multiple independent losses of *rpl20* and a single loss of *rps11* can be verified. However by reannotating published data and resequencing specimens when possible, we were able to identify the majority of genes that have been reported as lost or pseudogenes from Florideophyceae mitochondria.

Key words: Rhodophyta, mitochondria, gene loss, parasite, *atp8*, *rpl20*.

Introduction

Endosymbiotic events have had a profound impact on eukaryotic evolution (Lane and Archibald 2008; Keeling 2010; Koonin 2010; Zimorski et al. 2014; Martin et al. 2015). All eukaryotes [with one recent exception (Karnkowska et al. 2016)] possess a mitochondrion or mitochondrion-related organelle (MRO) that was initially acquired from an alpha-proteobacteria endosymbiont (Gray et al. 1999; Lang et al. 1999; Koonin 2010; Gray 2012). Additionally, photosynthetic lineages maintain a plastid that originated as a cyanobacterial endosymbiont in the shared ancestor of Glaucophytes, Rhodophytes, and Viridiplantae (Chlorophytes and Streptophytes), and was subsequently spread through the eukaryotic tree of life via secondary and tertiary endosymbiotic events (Bhattacharya et al. 2004; Keeling 2004; Stiller 2007; Gould et al. 2008; Lane and Archibald 2008; Keeling 2010; Stiller et al. 2014). There is evidence of massive gene transfer from the endosymbiont to the host nucleus upon the initial acquisition of these organelles, resulting in host control and regulation of the organelle's function (Martin et al. 1998; Timmis et al. 2004; Qiu et al. 2013; Ku, Nelson-Sathi, Roettger, Garg, et al. 2015; Ku, Nelson-Sathi, Roettger, Sousa, et al. 2015). Further organellar genome modifications appear to be mostly lineage specific, with gene losses and transfers being restricted within lineages (Tucker 2013; Janouškovec et al. 2013; Ku, Nelson-Sathi, Roettger, Garg, et al. 2015; Ku, Nelson-Sathi, Roettger, Sousa, et al. 2015; Qiu et al. 2015; Tanifuji et al.

© The Author(s) 2016. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://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

2016). The conservation among organellar genomes, in addition to their being inherited predominately uniparentally, has made organelles prime targets for understanding evolutionary relationships across and within the eukaryotic tree of life.

Red algae (phylum Rhodophyta) diversified from their last common ancestor, shared with green algae, more than 1 billion years ago (Yoon et al. 2004). There are ~7,100 currently described species of rhodophytes that are divided into seven classes; Bangiophyceae, Compsopogonophyceae, Cyanidiophyceae, Florideophyceae, Porpyridiophyceae, Rhodellophyceae, and Stylonematophyceae (Guiry MD and Guiry GM 2016). The Florideophyceae exhibit a wide range of morphological complexity and are by far the most species rich class, containing ~6,750 species spread across 30 orders (Guiry MD and Guiry GM 2016). Understanding the evolutionary relationships within the Florideophyceae has traditionally been complicated by phenotypic plasticity (Cianciola et al. 2010). More recently, molecular data have been analyzed and great progress has been made in describing new genera and species (Cianciola et al. 2010; Saunders and McDonald 2010; Le Gall and Saunders 2010). However, teasing apart the evolutionary histories of red algal orders has proven quite difficult even with the abundance of sequence data currently available (Verbruggen et al. 2010; Lam et al. 2016). Resolving the evolutionary relationships among florideophytes will provide a robust framework for asking a wide range of evolutionary questions including, but not limited to, transitions from marine to freshwater habitats, the evolution of the complex triphasic life-cycle found in many Florideophyceae orders, and the evolution of parasitism, a life strategy that has arisen many times among the Florideophyceae (Blouin and Lane 2012; Salomaki and Lane 2014; Blouin and Lane 2015; Lam et al. 2016). The use of the maternally inherited mitochondrial genome to resolve evolutionary relationships among the Florideophyceae shows promise (Yang et al. 2015).

The number of sequenced red algal organellar genomes has been increasing exponentially over the past decade. In part, this is a result decreasing sequencing costs allow for increasing use of next-generation sequencing technologies. Currently there are 30 published Florideophyceae mitochondrial genomes species available on GenBank (table 1). However, only 16 of the 30 florideophycean orders are represented in these data, and 10 of those orders are represented by a single mitochondrion genome sequence.

Analyses of mitochondrion and MRO genomes across the tree of life have shown they are highly variable in gene content, arrangement, and structure (Smith and Keeling 2015). More recently, the oxymonad Monocercomonoides was shown to have entirely lost its MRO and all genes of mitochondrial origin that had been transferred to the nucleus (Karnkowska et al. 2016). Wide variability of mitochondrial genome content and structure has been implicated in the Florideophyceae as well. A study investigating the impact of adopting a parasitic life strategy on mitochondrial genomes of red algae described the *atp8* and *sdhC* genes of red algal parasite *Gracilariophila oryzoides* as pseudogenes, and that the *atp8* gene in the parasite *Plocamiocolax pulvinata* has been lost entirely (Hancock et al. 2010). The authors concluded that the products of these genes may be provided either from the parasite nucleus as a result of endosymbiotic gene transfer (EGT), or perhaps the proteins are being obtained from their hosts. More recently, Yang et al. (2015) sequenced 11 Florideophyceae mitochondrial genomes. Analysis of their data, in combination with all previously sequenced red algal mitochondria led them to describe multiple independent losses of *atp8, nad4L, rpl20, rps11, sdhC, sdhD, secY*, and *ymf39* across the Florideophyceae (Yang et al. 2015).

Prior to this study, 30 Florideophyceae mitochondrial genomes have been sequenced. Of those, 19 are reported to be missing a functional copy of at least one gene. A total of eight different genes have been reported as pseudogenes or missing entirely from a Florideophyceae mitochondrial genome. Previous speculation on what is driving gene loss from Florideophyceae mitochondria include EGT from the mitochondrion to the nucleus (Hancock et al. 2010; Yang et al. 2015), and decreasing selective pressures in parasite mitochondria as a parasite may be obtaining products of those genes from the host (Hancock et al. 2010). Both explanations seem plausible, with the later hypothesis being directly responsible for the sequencing of the mitochondrial genome from the parasitic red alga, *Choreocolax polysiphoniae* and its host *Vertebrata lanosa* (this study).

The mitochondrial genomes of the parasitic red alga, *C. polysiphoniae* and its host *V. lanosa* represent the first mitochondrial genomes available from the family Rhodomelaceae, which comprises ~1/7th of species diversity within the phylum Rhodophyta (Guiry MD and Guiry GM 2016). In 2010, our lab reported that two mitochondrial respiratory protein-coding genes were degraded in the red algal parasites, *Gr. oryzoides* and *P. pulvinata* (Hancock et al. 2010). Unexpectedly, the *C. polysiphoniae* mitochondrion has no degradation of respiratory mitochondrial genes. To reconcile these datasets we resequenced the mitochondrial genomes of the parasite *Gr. oryzoides* and its host *G. andersonii*. Furthermore, we systematically reevaluate the described gene losses from the other 30 previously published Florideophyceae mitochondrial genomes, revealing that more than two-thirds of the described losses are the result of errors in sequencing or downstream analyses. We find Florideophyceae mitochondrial genomes to be highly conserved and that gene losses are rare and predominately, if not entirely, observed in genes encoding ribosomal proteins.

Table 1
Table of all currently available Florideophyceae mitochondrion genomes that were examined in this study with GenBank Accession, genome length, and A/T%

Species	GenBank Accession	Length	AT Content (%)	Reported Missing Genes	Notes
Ah. plicata	KF649303	32,878	66.6	*atp8, rpl20*	• ***atp8***: PCR of *atp8* region shows gene is present (KX687876). Previously published assembly was missing a single nucleotide, which resulted in a pseudogene. • ***rpl20***: With ATA start codon, gene is fully present (243 bp), but ATG start codon gene is short (147 bp). Translations of both align with other red algal *rpl20*.
A. taxiformis	KJ398158	26,097	73.3	*sdhD*	• ***sdhD***: Annotated in GenBank as 'gene/hypothetical protein CDS', BlastP of translated sequence hits other red algal *sdhD* sequences and aligns well with other translated red algal *sdhD* genes.
Ce. japonicum	KJ398159	26,200	71.5	*atp4 (as ymf39), sdhC, sdhD, TatC (as secY), rpl20*	• ***atp4***: Annotated in GenBank as 'gene/hypothetical protein CDS', BlastP of translated sequence hits other red algal *atp4* (ATP synthase B chain precursor) sequences and aligns well with other translated red algal *atp4* (*ymf39*) genes. In Yang et al. (2015), considered *atp4*, this is a case of nomenclature causing confusion. • ***sdhC***: In the correct location there is a region annotated as 'gene/hypothetical protein CDS', but it is shorter than other Florideophyceae *sdhC* genes. Translation shows conservation of residues, particularly at 5′ region of the sequence. Possibly truncated at 3′ end, resequencing would help clarify. • ***sdhD***: Found in published data, but was unannotated 16,636>16,403. • ***TatC***: Seems to be the result of homopolymer sequence error, though possibility of pseudogene remains. Homology based on translation alignment of region from 23,168>23,590. • ***rpl20***: Likely an actual loss, all sequenced members of Ceramiales also missing *rpl20*. Truncated from 5′ end and about half the gene remains as a pseudogene.
Ce. sungminbooi	KU145004 and KU145005	24,508 24,494	71.2 71.2	noted as partial in article (Hughey and Boo 2016)	• ***rpl20***: Likely a loss, all sequenced members of Ceramiales also missing *rpl20*. Truncated from 5′ end and about half the gene remains as a pseudogene. • ***sdhC***: Found, as unannotated 10,977 > 10,678 in conserved location between *atp*9 and *sdh*B in KU145004. • ***sdhD***: Found, as unannotated ORF 15,284 > 15,036 in conserved location between *nad*4 and *nad*2 in KU145004. • ***TatC***: Found, as unannotated ORF 15,277 > 15,029 in conserved location between *nad*4 and *nad*2 in KU145004.
Ch. crispus	NC_001677	25,836	72.1	—	• Complete
C. polysiphoniae	KX687877	25,357	79.4	*rpl20*	• ***rpl20***: Likely an actual loss, all sequenced members of Ceramiales also missing *rpl20*. Truncated from 5′ end and about half the gene remains as a pseudogene.
Co. compressa	KU053956	25,391	74.3	—	• ***rpl20***: Truncated and here considered a pseudogene.
Corallina officinalis	KU641510	26,504	69.9		• Although the manuscript is published, not yet available on GenBank • Reported use of GTG (*nad*2) and ATT (*sdh*C) as start codons, otherwise reported as complete
D. binghamiae	KX247283	26,052	77.4		• ***sdhC***: Found, as unannotated 21,403 > 21,789 in conserved location between *atp*9 and *sdh*B. • ***rpl20***: No good evidence of *rpl20* homology • ***rps3***: Truncated version published seems to be the result of homopolymer sequence error, though possibility of pseudogene remains. Insertion of a nucleotide around 23,590 restores conservation of start and end of gene compared with other red algal *rps3* genes

(continued)

Table 1 Continued

Species	GenBank Accession	Length	AT Content (%)	Reported Missing Genes	Notes
					• *rps11*: Moving the initiation codon from 12,194 to 12,044 removes overlap with 3′ of the *nad3* gene and results in conserved start with other Florideophyceae *rps11* genes based on alignment.
G. elegans	KF290995	24,922	70.5	*rpl20*	• *rpl20*: With ATA start codon, gene is present but short (177 bp) with some conserved residues. Absent with ATG start. Area somewhat conserved but appears to be degraded too much to actually encode *rpl20*.
G. vagum	KC875854	24,901	69.5	*rpl20*	• *rpl20*: With ATA start codon, gene is mostly present (222 bp) with some conserved residues. With ATG start it could be either 297 bp with overlapping tRNAs or short at 186 bp. Still considered a pseudogene here.
Gracilaria chilensis	KP728466	26,898	72.4	—	• Complete
Gracilaria salicornia	KF852534	25,272	71.6	—	• Complete
G. vermiculophylla	KJ526627	25,973	71.9	—	• Complete
Gr. oryzoides	NC_014771 and KX687879	25,161	71.9	*atp8, sdhC*	• *atp8*: Present and complete, early stop codon assumed to be sequencing error. • *sdhC*: Present and complete, early stop codon assumed to be sequencing error. • **Newly sequenced *Gr. oryzoides* mitochondrion available at GenBank (KX687879).**
G. andersonii	NC_014772 and KX687878	27,036	72	*atp4 (as ymf39)*	• *atp4*: Present in resequenced *G. andersonii* mitochondrion. • *rps11*: Originally reported as inversion which appears to result from frameshift mutation in original sequence. This is corrected in resequenced mitochondrion. • **Newly sequenced *G. andersonii* mitochondrion available at GenBank (KX687878).**
G. chorda	NC_023251	26,534	72.4	*atp4 (as ymf39)*	• atp4: Annotated as gene/hypothetical protein CDS in GenBank but considered as "present" in Yang et al. (2015). Definitely present.
Gracilariopsis lemaneiformis	JQ071938	25,883	72.5	—	• Complete
Gra. angusta	NC_023094	27,943	69.8	—	• Complete, but uses multiple start codons, seemingly where unnecessary (see table 4). Contains hypothetical protein CDS in *cox1* intron.
Grateloupia taiwanensis	KM999231	28,906	68.6	—	• Complete. Contains hypothetical protein CDS in *cox1* intron that is also annotated as *cox1*. • ***atp4***: Initiation codon questionable, see table 4.
H. rubra	KF649304	33,066	67.8	*atp4, atp8, rpl20*	• *atp4*: Found, as unannotated 13,225 > 13785. ORF was at conserved location immediately after *cox3*. Translation conserved with other Florideophyceae *atp4* genes. • *atp8*: Found, as unannotated 24,587 > 24,213. ORF present in conserved location between *atp8* and *nad5*, 375 bp and conserved residues at n-terminus. • *rpl20*: Truncated 5′ region–135 bp, here considered a pseudogene.
K. striatus	KF833365	25,242	69.9	*atp4 (as ymf39)*	• *atp4*: Annotated as *atp4* gene. This is a case of nomenclature causing confusion.
Mastocarpus papillatus	KX525587	25,906	65.0		• Complete
P. palmata	KF649305	29,735	67.8	—	• Complete
P. pulvinata	HQ586061	25,894	76.1	*atp8, nad4L*	• *atp8*: Found, as annotated gene/hypothetical protein CDS in GenBank. 20,389 > 19,985. • *nad4L*: Found, as unannotated 25,592 > 25,894.

(continued)

Table 1 Continued

Species	GenBank Accession	Length	AT Content (%)	Reported Missing Genes	Notes
					• *atp4*: Moving the initiation codon from 7,772 to 7,790 will remove overlap with 3′ of the *cox3* gene and result in conserved start with other Florideophyceae *atp4* genes based on alignment.
Pl. cartilagineum	KJ398160	26,431	76.4	*atp8, nad4L*	• *atp8*: Present, as annotated gene/ATP synthase F0 subunit 8 CDS in GenBank. 20,528 > 20,127. • *nad4L*: Found, as unannotated 26,172 > 43 (linear sequence of circular molecule ends and starts over from the beginning of sequence). • *atp4*: Moving the initiation codon from 7,766 to 7,784 will remove overlap with 3′ of the *cox3* gene and result in conserved start with other Florideophyceae *atp4* genes based on alignment.
R. pseudopalmata	KC875852	26,351	70.5	*rpl20*	• *rpl20*: ATG start codon leaves it 40 residues shorter than other, while ATT start codon leaves it four residues shorter than others. Here considered present with ATT start. RNA for this gene would be very useful to confirm whether this is transcribed and not a pseudogene.
Riquetophycus sp.	KJ398161	26,351	74.3	—	• Complete
S. schousboei	KJ398162	25,906	73.3	*rpl20*	• *rpl20*: Found, as unannotated 23,912 > 24,148.
Sc. dubyi	KJ398163	26,438	74.1	*atp4 (as ymf39), rpl20*	• *atp4*: Annotated in GenBank as "gene/hypothetical protein CDS", BlastP of translated sequence hits other red algal *atp4* and ATP synthase B chain precursor sequences and aligns well with other translated red algal *atp4* genes. In Yang et al. (2015), considered *atp4*. • *rpl20*: Annotated in GenBank as 'gene/hypothetical protein CDS', BlastP of translated sequence shows similarity to other red algal *rpl20* genes.
S. flabellata	KJ398164	26,767	71.5	*atp4 (as ymf39), sdhC*	• *atp4*: Annotated in GenBank as "gene/hypothetical protein CDS", BlastP of translated sequence hits other red algal atp4 and ATP synthase B chain precursor sequences and aligns well with other translated red algal *atp4* genes. In Yang et al. (2015), considered *atp4*. • *sdhC*: Found a conserved ORF that relies on a TTA start codon (11,358 > 10,966).
Sp. durum	KF186230	26,202	71.6	*rps11, rpl20*	• *rpl20*: Many options exist for start codons other than ATG, clearly conserved residues present. Considered present in this study using ATT start codon table 4. • *rps11*: Pseudogene or sequencing error, regional conservation remains.
V. lanosa	KX687880	25,119	71.7	*rpl20*	• *rpl20*: Likely an actual loss, all sequenced members of Ceramiales also missing *rpl20*. Truncated from 5′ end and about half the gene remains as a pseudogene.

Note.—Genes previously reported as missing are listed along with notes regarding their status as a result of this study.

Materials and Methods

Mitochondrial Genome Sequencing

Specimens of *V. lanosa* and *C. polysiphoniae* were collected from Beavertail State Park, Jamestown, Rhode Island, USA (voucher RI 0423). *Gracilariopsis andersonii* and *Gr. oryzoides* were collected at Pigeon Point, Pescadero, California, USA (voucher CL031613). Representatives of these parasite and host pair populations are retained as vouchers in the Lane Lab herbarium at the University of Rhode Island. Vegetative tissue from *V. lanosa* and *G. andersonii* was inspected for parasites and epiphytes under a dissecting microscope and clean tissue was ground under liquid nitrogen. Erumpent pustules of *C. polysiphoniae* (n = 50) and *Gr. oryzoides* (n = 10) were excised from the thallus of their hosts *V. lanosa*, and *G. andersonii*, respectively, and collected in a 1.5 ml microcentrifuge tube. The parasite tissue was hand-ground using a Corning Axygen PES-15-B-SI disposable tissue grinder pestle in a 1.5 ml microcentrifuge tube while submerged in 100 μl of DNA extraction buffer (Saunders 1993). DNA was extracted from all specimens using a standard phenol/chloroform extraction (Saunders 1993).

All DNA libraries were prepared for Illumina sequencing on the Apollo 324 robot using the PrepX ILM DNA Library Kit

(Wafergen Biosystems, Freemont, California). The *G. andersonii* library was sequenced on a full-cell of an Illumina MiSeq paired-end 250 × 250 basepair (bp) run yielding 30,330,114 sequences in pairs. The *Gr. oryzoides* and *C. polysiphoniae* libraries were each sequenced on full-cells of an Illumina MiSeq paired-end 300 × 300 bp run yielding 26,097,992 and 29,355,470 sequences in pairs, respectively. The *V. lanosa* library was sequenced on a partial-cell of an Illumina MiSeq paired-end 300 × 300 bp run yielding 12,888,082 sequences in pairs. For all datasets, sequences with PHRED scores <30 were removed and the remaining reads were trimmed of adapter sequences. Additionally, fifteen 5′ and five 3′ nucleotides were trimmed from the remaining reads and all reads under 100 nucleotides were removed from the dataset. All trimming was completed using CLC Genomics Workbench v. 8.5.1 (CLC Bio-Qiagen, Aarhus, Denmark) and the remaining reads were assembled using default parameters in CLC Genomics Workbench v. 8.5.1. Trimmed sequencing reads for the *G. andersonii* and *Gr. oryzoides* mitochondria were mapped back to the previously published mitochondrion to compare the two assemblies and confirm support for differences.

Open reading frame (ORF) prediction on the *V. lanosa* and *C. polysiphoniae* mitochondrion sequences was done using translation table 4 (Protazoa Mitochondrion) using ATG as a start-codon in Geneious Pro v9.1 (Kearse et al. 2012). The resulting ORFs were manually annotated using blastN against GenBank. If blastN was insufficient for annotating an ORF, the ORF was translated to an amino acid sequence and then searched against the non-redundant protein sequence database (nr) in GenBank using blastP and the Pfam database (Finn et al. 2010, 2015). If no conserved domain or sequence similarity could be found after searches using blastP or Pfam, the ORF remained without further annotation. Mitochondrion genome sequences were submitted to the tRNAscan-SE online server v1.21 for identification of tRNA sequences (Schattner et al. 2005). Ribosomal RNA predictions were based on annotations produced by MFannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl).

Red Algal Mitochondrial Genome Conservation

All 30 currently available Florideophyceae mitochondrial genomes (table 1) were downloaded from GenBank and imported into GeneiousPro v9.1. These mitochondrial genomes were combined with those from *V. lanosa* and *C. polysiphoniae* to create a database of Florideophyceae mitochondrial genomes. Sequences that have previously been found to have missing genes or pseudogenes were reanalyzed for ORFs in GeneiousPro v9.1. In cases where an ORF was found in a conserved location that had not previously been annotated as a gene, the ORF was translated and searched against the non-redundant protein sequence database (nr) in GenBank using blastP and Pfam. If this was insufficient to annotate an ORF in a conserved location, representatives of the missing genes were mapped back to the genome of interest for further evaluation and the region was manually curated. Translations of ORFs from locations of missing genes were aligned with annotated copies of those genes to manually assess annotation. When an apparent premature stop codon was found in a conserved location for a missing gene or pseudogene, the region was resequenced using PCR amplification for confirmation when material or DNA from that species could be obtained.

To determine the AT content (%) and non-synonymous to synonymous substitution ratio (d*N*/d*S* ratio) (table 2), all protein-coding genes widely shared throughout the 32 Florideophyceae mitochondrial genomes were aligned using GeneiousPro v9.1. The average AT content (%) was calculated for each gene across the Florideophyceae in GeneiousPro v9.1. The CODEML program in PAML v. 4.8 (Yang 2007) was utilized to estimate the pairwise d*N*/d*S* ratio across all published Florideophyceae mitochondrion genes. For each gene analyzed, a nucleotide alignment was created using the translation align function in GeneiousPro v9.1 utilizing the Blosum62 cost matrix. Additionally, a Neighbor-Joining tree was constructed from the alignment in GeneiousPro v9.1 using the Tamura-Nei substitution model and the gene from *Hildenbrandia rubra* was used to root the tree. For the *rpl*20 gene, which has been lost in *H. rubra*, *Palmaria palmata* was used to root the tree. The alignment and Neighbor-Joining tree were used as input files and the following parameters were specified in CODEML: runmode = 0; seqtype = 1; codonfreq = 0; model = 0; icode = 4; and omega (measures d*N*/d*S* ratio) and kappa (measures transitions/transverstions) were estimated.

Results and Discussion

Description of a Red Algal Alloparasite and Host Mitochondrion

The mitochondrial genomes of the red algal alloparasite, *C. polysiphoniae* (KX687877) and its host, *V. lanosa* (KX687880) represent the first mitochondrial genomes sequenced from the Rhodomelaceae (Florideophyceae, Rhodophyta), further expanding the diversity of available red algal mitochondrial genomes. The *C. polysiphoniae* mitochondrial genome is a 25,357 bp long circular molecule with an AT content of 79.4%. The mitochondrial genome of *C. polysiphoniae* encodes 23 protein-coding genes, 2 rRNAs and 20 tRNAs and contains only 9.9% intergenic, non-coding DNA. The *V. lanosa* mitochondrial genome is also a circular molecule that is 25,337 bp long, has an AT content of 76.4%, and encodes 23 protein-coding genes, 2 rRNAs and 19 tRNAs with 10.4% of the mitochondrial DNA being non-coding. Both the *C. polysiphoniae* and *V. lanosa* mitochondrial genomes maintain a

Table 2
A/T%, non-synonymous to synonymous mutation (d*N*/d*S*) ratio and individual d*N* and d*S* values for genes encoded on the Florideophyceae mitochondrion genomes

Gene	AT Content (%)	d*N*/d*S* ratio	d*N*	d*S*	Species with pseudogene only
atp4	79.8	0.51645	022882	0.44307	
atp6	72.5	0.12864	0.07486	0.58195	
atp8	78.8	0.51248	0.20110	0.39241	
atp9	66.1	0.01550	0.00212	0.13673	
cob	70.3	0.11691	0.05443	0.46555	
cox1[a]	67.2	0.06936	0.01711	0.24675	
cox2	69.5	0.14959	0.04549	0.30414	
cox3	68.3	0.12819	0.07186	0.56058	
nad1	69.0	0.09454	0.06667	0.70520	
nad2	74.6	0.31481	0.15877	0.50434	
nad3[b]	74.5	0.15718	0.07060	0.44919	
nad4	72.7	0.18548	0.07790	0.41999	
nad4L	75.3	0.14619	0.05654	0.38677	
nad5	71.9	0.21716	0.08547	0.39356	
nad6	75.1	0.26439	0.17651	0.66762	
rpl16	75.9	0.36193	0.13664	0.37752	
rpl20	79.1	0.62160	0.46012	0.74023	*Ce. japonicum*
					Ce. sungminbooi
					C. polysiphoniae
					Co. compressa
					D. binghamiae[d]
					G. elegans
					G. vagum
					H. rubra
					V. lanosa
rps3[c]	76.9	0.51571	0.28532	0.55326	
rps11	77.3	0.45177	0.19375	0.42886	*Sp. durum*
rps12	67.5	0.20794	0.11986	0.57642	
sdhB	71.7	0.20603	0.10707	0.51969	
sdhC	78.6	0.47084	0.17487	0.37140	
sdhD	78.2	0.46556	0.26263	0.56412	
TatC	80.0	0.50046	0.34550	0.69037	*Ce. japonicum*

Note.—Species with a pseudogene, rather than a functional copy of the gene, are listed in the far right column and were left out of calculations of A/T% and d*N*/d*S* ratio.
[a]Introns removed and CDSs only were used for A/T% and d*N*/d*S* analysis.
[b]*Ce. japonicum nad3* left out of d*N*/d*S* analysis.
[c]*D. binghamiae rps3* left out of d*N*/d*S* analysis.
[d]No evidence for remnant pseudogene, appears to be a complete loss.

similar genome architecture to other published red algal mitochondria.

Parasitic Red Algal Mitochondrial Genomes Are Conserved

The *atp8* gene has previously been reported missing from five Florideophyceae mitochondria (Hancock et al. 2010; Yang et al. 2015) including the parasites *Gr. oryzoides* and *P. pulvinata*. A re-annotation of the mitochondrion of *P. pulvinata*, identified the *atp8* as an ORF that was annotated only as a hypothetical protein coding sequence (CDS) in the sequence downloaded from GenBank (tables 1 and 3). Subsequently, resequencing the *Gr. oryzoides* mitochondrial genome (KX687879) revealed a complete copy of the *atp8* gene, rather than the pseudogene that was previously reported using both Illumina and Sanger sequencing (Hancock et al. 2010). The *sdhC* gene was also previously reported to be a pseudogene in the adelphoparasite *Gr. oryzoides* (Hancock et al. 2010). As with *atp8*, resequencing of the *Gr. oryzoides* mitochondrion (KX687879) demonstrated that there was no frameshift mutation, as originally published, and that *sdhC* remains complete in red algal parasite mitochondria. These findings indicate that red algal parasites have not found alternative mechanisms for acquiring mitochondrion proteins and rely on their own mitochondrion for generating cellular energy as was previously hypothesized (Hancock et al. 2010).

Gene Loss in Other Red Algal Mitochondria

As a result of identifying conserved copies of genes originally reported to have been lost, we reevaluated all reported gene losses from red algal mitochondria. Investigation of the other reported *atp8* losses revealed that the gene was an ORF annotated as a hypothetical protein CDS in *Plocamium cartilagineum*, and that an ORF corresponding to *atp8* could be found in the published *H. rubra* sequence data that were not previously annotated (tables 1 and 3). The *Ahnfeltia plicata* mitochondrion had a premature stop codon resulting in a pseudogene where *atp8* is normally found, however targeted PCR and sequencing showed the gene (KX687876) is complete. Analysis of the ratio of non-synonymous to synonymous substitutions (d*N*/d*S* ratio) in Florideophyceae copies of the *atp8* gene show a higher rate on non-synonymous mutations in *atp8* than *atp6* and *atp9*, which combine with *atp8* to make up the F_0 domain of the F_1F_0-ATP synthase complex involved in ATP synthesis (table 2). However the d*N*/d*S* ratio of all three proteins remains <1 indicating that purifying selection is acting on the mitochondrial F_1F_0-ATP synthase complex in red algae, as is expected from genes essential for mitochondrion function.

Although the biological implications of losing the *nad4L* gene was not discussed in previous literature, the gene was noted as being absent in the mitochondrial genomes of both *P. pulvinata* and *Pl. cartilagineum* (Hancock et al. 2010; Yang et al. 2015). In *P. pulvinata* an ORF was identified in the same location as other red algal copies of *nad4L*, between the 16S ribosomal RNA and the 26S ribosomal RNA, and both Pfam sequence search and blastP search of the translation strongly supports it coding for a functional *nad4L*. The published sequence for the mitochondrial genome of *Pl. cartilagineum* splits an ORF here identified as the *nad4L* gene in two pieces, with the 5′ portion of the sequence found from bases 26,172–26,431 and the 3′ portion of the sequence is located from bases 1 to 43. With a *dN*/*dS* ratio of 0.14619, the *nad4L* gene remains under strong purifying selection in red algal mitochondria. Therefore, the loss of *nad4L* in any red algal mitochondria would represent a strong departure from this heavy selective pressure.

The *sdhD* gene encodes an essential protein that serves to anchor the succinate dehydrogenase complex II to the innermembrane of the mitochondrion (Elorza et al. 2004; Bayley et al. 2005, 2006). The *sdhD* gene was reported missing from the mitochondria of *Ceramium japonicum* and *Asparagopsis taxiformis* (Yang et al. 2015) and the gene is also not annotated in the more recently published mitochondrial genome of *Ceramium sungminbooi* (Hughey and Boo 2016). Upon reanalysis, an ORF was identified between *nad4* and *nad2*, in the conserved Florideophyceae location of *sdhD* in the published mitochondrial genomes for all three species (see tables 1 and 3). Furthermore, a translated alignment of these ORFs with other Florideophyceae copies of *sdhD* show they are conserved in frame, retaining several critical conserved residues (fig. 1), and therefore should be annotated as *sdhD*.

Table 3

Current status of Florideophyceae mitochondrial genes previously reported as missing in Hancock et al. (2010) and Yang et al. (2015) or otherwise unannotated

Gene	Current Status		
	Gene Present	Location in Published Sequence or GenBank Accession Number	Pseudogene
atp8	*Ah. plicata*	KX687876	
	Gr. oryzoides	20,431 > 20,024[d]	
	H. rubra	24,587 > 24,213	
	P. pulvinata	20,389 > 19,985	
	Pl. cartilagineum	20,528 > 20,127	
nad4L	*P. pulvinata*	25,592 > 25,894	
	Pl. cartilagineum	26,129 > 26,474 (43)	
rpl20	*Ah. plicata*[a]	30,851 > 31,093	*Ce. japonicum*
	R. pseudopalmata[a]	24,127 > 24,351	*Ce. sungminbooi*
	S. schousboei	23.912 > 24,148	*C. polysiphoniae*
	Sc. dubyi	24,440 > 24,709	*Co. compressa*
	Sp. durum[a]	24,248 > 24,487	*D. binghamiae*[f]
			G. elegans
			G. vagum
			H. rubra
			V. lanosa
rps11			*Sp. durum*
sdhC	*Ce. japonicum*	12,331 > 12,044	
	Ce. sungminbooi	10,977 > 10,678	
	D. binghamiae	21,403 > 21,789	
	Gr. oryzoides	11,593 > 11,234[d]	
	S. flabellata[a]	11,358 > 10,966	
sdhD	*A. taxiformis*	15,759 > 15,514	
	Ce. japonicum	16,636 > 16,403	
	Ce. sungminbooi	15,284 > 15,036	
TatC	*Ce. sungminbooi*	15,277 > 15,029	*Ce. japonicum*[g]
atp4	*Ce. japonicum*[b]	6,935 > 7,489	
	G. andersonii	7,560 > 8,102[e]	
	G. chorda[c]	7,238 > 7,780	
	H. rubra	13,225 > 13,785	
	Sc. dubyi[b]	7,261 > 7,803	
	S. flabellata[b]	7,234 > 7,776	

[a]Indicates presence of functional gene is dependent on non-ATG start codon, alternatively these could be pseudogenes. RNA sequence data would be required to confirm gene function.

[b]The *atp4* gene was annotated as hypothetical protein CDS in GenBank but considered as *atp4* in Yang et al. (2015), figure 2.

[c]The *atp4* gene was annotated as hypothetical protein CDS in GenBank but considered as *ymf39* in Yang et al. (2015), figure 2.

[d]Location in newly sequenced *Gr. oryzoides* mitochondrion (GenBank KX687879).

[e]Location in newly sequenced *G. andersonii* mitochondrion (GenBank KX687878).

[f]No evidence for remnant pseudogene.

[g]The *Ce. japonicum TatC* gene seems likely to be the result of homopolymer sequence error, though possibility of pseudogene remains. Due to high levels of variation in length and sequence of Florideophyceae *TatC* genes, we continue to recognize the *Ce. japonicum TatC* gene as a pseudogene until firm evidence contradicts this.

Table 4
The use of alternative start codons by gene based on published literature and the proximity to the closest in-frame ATG initiation codon. Alternative initiation codons that are supported by the lack of a nearby ATG initiation codon and conserved gene start location based on alignment are indicated in bold

Gene	Species	Location with Alternative Initiation Codon	Location with ATG Initiation Codon	Difference in Gene Length in nucleotides (Amino Acids)
atp4	*A. taxiformis*	7,298 (ATT)	7,304	6 (2)
atp6	*Gra. angusta*	23,475 (ATT)	23,466	9 (3)
	Sp. durum	21,462 (ATT)	21,465	3 (1)
atp8	—	—	—	—
atp9	—	—	—	—
cob	***S. flabellata*[a]**	**8,070 (ATT)**	**8,196**	**126 (42)**
cox1	*Sc. dubyi*	3,889 (ATT)	3,904	15 (5)
cox2	*Gra. angusta*	7,845 (TTG)	7,818	27 (9)
	P. palmata	6,556 (ATT)	6,535	21 (7)
cox3	—	—	—	—
nad1	—	—	—	—
nad2	—	—	—	—
nad3	—	—	—	—
nad4	*H. rubra*	22,165	22,150	15 (5)
nad4L	—	—	—	—
nad5	—	—	—	—
nad6	—	—	—	—
rpl16	***Ah. plicata*[a]**	**5,899 (TTA)**	**6,169**	**270 (90)**
	***Gra. angusta*[a]**	**3,324 (TTA)**	**3,585**	**261 (87)**
rpl20	***Ah. plicata*[a,c]**	**30,851 (ATA)**	**30,947**	**96 (32)**
	***R. pseudopalmata*[a]**	**24,127 (ATT)**	**24,235**	**108 (36)**
	***Sp. durum*[a]**	**24,248 (ATT)**	—	—
rps3	*Pl. cartilagineum*	2,693 (ATC)	2,732	39 (13)
	Sp. durum	2,574 (ATA)	—	—
rps11	***Ah. plicata*[a]**	**17,436 (ATA)**	**17,277**	**159 (53)**
	G. chilensis	10,426 (ATT)	10,423	3 (1)
	Gra. angusta	14,946 (ATT)	14,938	9 (3)
	***P. palmata*[a]**	**14,083 (TTA)**	—	—
rps12	—	—	—	—
sdhB	***Gra. angusta*[a]**	**13,459 (TTA)**	**13,312**	**147 (49)**
sdhC	*Gra. angusta*	13,852 (TTG)	13,849	3 (1)
	H. rubra[a,b]	17,061 (ATT)	17,022	39 (13)
	Sc. dubyi	11,319 (CTT)	11,310	9 (3)
sdhD	*Ah. plicata*	20,718 (TTA)	20,733	15 (5)
	H. rubra[a]	20,662 (ATA)	—	—
TatC	*Ah. plicata*	29,469 (ATT)	29,487	18 (6)
	***Ch. crispus*[a,d]**	**348 (GTT)**	**681**	**333 (111)**
	***Gra. angusta*[a]**	**24,671 (ATT)**	**25,031**	**360 (120)**
	***H. rubra*[a]**	**29,085 (ATA)**	—	—
	***K. striatus*[a]**	**24,437 (TTG)**	**24,176**	**261 (87)**
	***P. palmata*[a]**	**24,152 (ATC)**	**23,687**	**465 (155)**
	***R. pseudopalmata*[a]**	**22,834 (TTA)**	**23,350**	**516 (172)**
	***Schimmelmania schousboei*[a]**	**22,636 (ATT)**	**23,188**	**552 (184)**
	S. flabellata[a]	23,492 (TTA)	23,546	54 (18)
	Sp. durum	22,915 (ATC)	22,942	27 (9)

[a]Indicates examples where other non-ATG initiation codons from translation table 4 (Protozoa Mitochondrion) are also possible locations for the gene to start although no ATG codon is found within 30 nucleotides (10 amino acid residues) upstream or downstream from the start of the currently annotated gene.

[b]The *H. rubra sdhC* gene annotation is longer than other copies of *sdhC* and the beginning of the gene overlaps with a tRNA. Starting annotation at ATG makes the gene much more similar in length to other Florideophyceae copies of *sdhC*.

[c]Gene not previously annotated in GenBank.

[d]The *Ch. crispus TatC* (*ymf16*) gene is currently annotated with a GTT initiation codon, which is not found for any other Florideophyceae mitochondrion gene nor is it a start codon in translation table 4 (Protozoa Mitochondrion). Four other ORFs in the same reading frame that use either ATA or TTA as a start codon for *TatC* gene are found from 12 to 39 nucleotides downstream of the GTT codon.

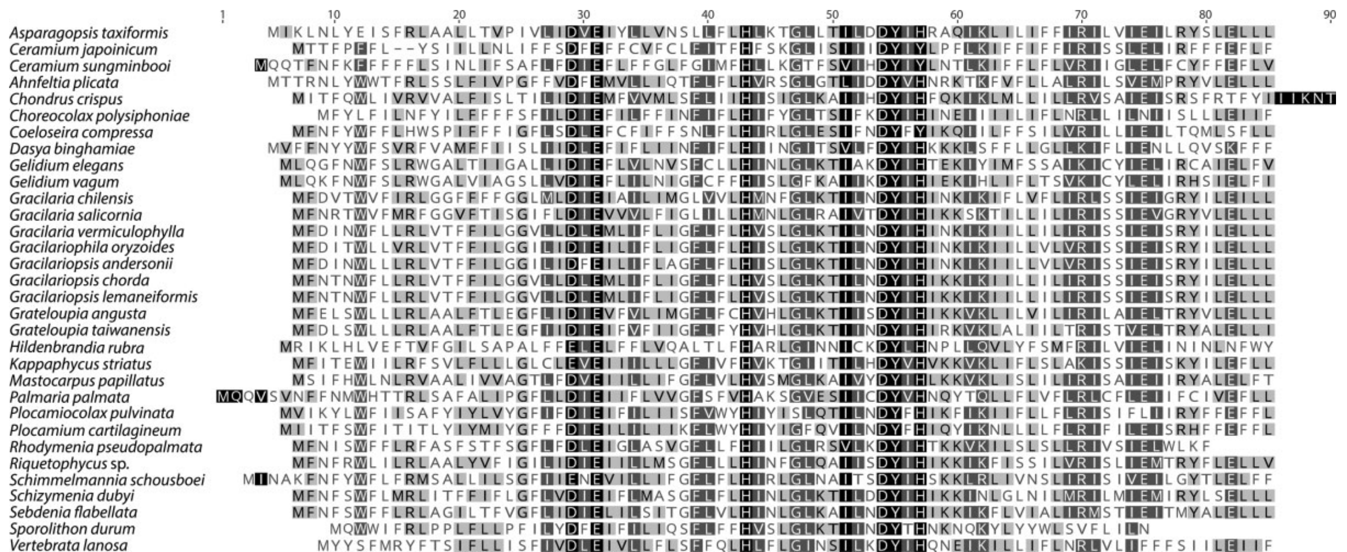

Fig. 1.—Translated alignment of *sdhD* genes from florideophycean mitochondria showing *A. taxiformis, Ce. japonicum*, and *Ce. sungminbooi* (top three sequences) share critical conserved residues with all other Florideophyceae *sdhD* genes.

Four mitochondria are reportedly missing copies of *sdhC*. Similarly to our findings with the *sdhD* genes, unannotated ORFs that are conserved with other Florideophyceae *sdhC* genes were identified from the mitochondrial genomes of *Ce. sungminbooi* and *Dasya bingamiae* (table 3). Based on the published *Sebdenia flabellata* mitochondrial genome, using an ATG as the only start-codon, there is no ORF that can be attributed to *sdhC*. However, using all start-codons in translation table 4 (Protozoa Mitochondrion) an ORF that is highly conserved in comparison with other Florideophyceae copies of *sdhC* is found with a TTA start-codon (tables 1 and 3). Alternative start codons have previously been invoked for annotating red algal mitochondrion genes with variable support, which is discussed in more detail below (and see table 4). The *S. flabellata sdhC* appears to be a well-justified case for using an alternative initiation codon.

The *Ce. japonicum* mitochondrion is the other reported case of an *sdhC* gene loss (Yang et al. 2015). Although it appears to be highly conserved throughout the 5′ region in comparison to other species, the *Ce. japonicum sdhC* is truncated by ~81 nucleotides (27 amino acids) at the 3′ end when aligned with copies of the *sdhC* gene from other Florideophyceae. The *Coeloseira compressa sdhC* is similarly conserved at the 5′ region and truncated at the 3′ end. A Pfam search of the *Ce. japonicum* and *Co. compressa* sequences, translated to amino acids, confirms their identity as Succinate dehydrogenase/Fumarate reductase transmembrane subunit proteins though suggests they may be truncated as well. Although material was not available for experimental validation, we speculate that this observed truncation has little effect on the functionality of *sdhC* as an anchor protein in succinate dehydrogenase complex II. The length of Florideophyceae *sdhC* genes (excluding *Ce. japonicum* and *Co. compressa*) is quite variable, ranging from 339 bp in *P. pulvinata* to 411 bp in *A. taxiformis*. Furthermore, the dN/dS ratio remains at 0.47084 indicating that purifying selection is acting fairly strongly on deleterious mutations in *sdhC*. The alternative would seem that the *sdhC* gene in *Ce. japonicum* and *Co. compressa* is losing its functional capacity, which would hinder the ability of these free-living species to generate cellular energy.

Although not reported as a loss, the published *G. andersonii rps11* gene is inverted in comparison with all other Florideophyceae copies of the gene (Hancock et al. 2010; Yang et al. 2015). Resequencing this genome revealed an ORF in the conserved location between *nad3* and *atp9* that was not inverted and maintained strong homology with red algal *rps11* genes. Analysis of this ORF in comparison to the previously published *G. andersonii* mitochondrion identified a string of seven “A”s stretching from bases 9,158 to 9,164 correspond to only six “A”s in the newly sequenced mitochondrion. This apparent frameshift mutation resulted in a premature stop codon in the conserved direction that led to identifying an ORF in the same location but inverted as *rps11* in the earlier publication. The *rps11* gene in the resequenced *G. andersonii* mitochondrion, extending from bases 14,568 to 14,209, maintains strong homology with, and is encoded in the same direction as other Florideophyceae copies of *rps11*.

Although no genes were explicitly described as being lost in the recently sequenced *D. binghamiae* mitochondrial genome (Tamayo and Hughey 2016), annotations for *rpl20* and *sdhC* are absent from the published sequence. Additionally, alignments demonstrate that the *cox3, rps3* and *TatC* genes are truncated in comparison with other Florideophyceae. Perhaps

even more interesting is the report of two inverted multi-gene rearrangements that are unprecedented in light of the highly conserved synteny in florideophycean mitochondria. Unfortunately a thorough evaluation of the losses, truncations and rearrangements in this mitochondrial genome is difficult as the publication is extremely brief (<500 words) and lacks essential details such as the sequencing platform from the materials and methods.

Frameshift Mutations Are Overstated

In addition to the annotation of an inverted *rps11* in *G. andersonii*, frameshift mutations have been described as the cause for genes being lost or becoming pseudogenes in Florideophyceae mitochondria including *atp8* in *Ah. plicata* and *Gr. oryzoides* and *sdhC* in *Gr. oryzoides*. The *Gracilaropsis andersonii rps12* gene is another case of an apparent frameshift mutation causing a gene to be truncated. In *G. andersonii*, the *rps12* gene is annotated at 240 nucleotides in length while other red algal copies of the gene range from 366 to 390 bp long. As a part of this study we resequenced the *G. andersonii* mitochondrion (KX687878) and identified that the "CT" found at bases 25,864–25,865 in the previously published *G. andersonii* mitochondrion appears to be the result of sequencing or assembly error. Without these additional bases, the *rps12* gene remains conserved and is 369 bp long.

At first glance, the *Ce. japonicum nad3* gene appears to be an instance of a frameshift mutation causing a gene to be truncated. Although the *Ce. japonicum nad3* gene is annotated in its mitochondrial genome, it is only 234 bp long, whereas all other Florideophyceae *nad3* genes are either 363 or 366 bp long. An alignment of other Florideophyceae *nad3* genes to the *Ce. japonicum nad3* region indicates that this truncation is the result of a frameshift mutation in a string of 26 "T" and 3 "C" between 32 and 60 bp from the start codon. A translated alignment of the annotated *Ce. japonicum nad3* with all other Florideophyceae *nad3* genes shows little conservation in the annotated *Ce. japonicum nad3*. Additionally, a blastP search of the NCBI nr database, and a Pfam sequence search of the translated original annotation shows the region is not homologous with any gene sequenced to date. However, manually deleting a "T" from the previously mentioned string yields a 366 bp *nad3* gene that is highly conserved with copies of the *nad3* gene sequenced from other Ceramiales mitochondria (fig. 2) and is homologous with *nad3* genes in the NCBI nr database and Pfam database. Long homopolymer runs are notoriously challenging for both sequencing and assembly (Kieleczawa 2006; Gilles et al. 2011; Loman et al. 2012; Laehnemann et al. 2016) but is a more likely explanation than a frameshift resulting in two conserved sections of the gene.

The *Ce. japonicum TatC* (*secY*) initially appears to be another case of a Florideophyceae mitochondrion gene losing function and becoming a pseudogene due to a frameshift mutation, and again, pinpointing the exact location of the mutation is difficult. By manually manipulating the sequence and deleting a nucleotide from a sting of 43 T's and 7 C's between 23,501 and 23,550 bp into the published sequence, an ORF that is highly conserved with other Florideophyceae *TatC* genes containing an ATT initiation codon is observed. Due to high levels of variation in length and sequence of Florideophyceae *TatC* genes, we continue to recognize the *Ce. japonicum TatC* gene as a pseudogene until firm evidence contradicts this. However, based on our findings that all frameshift mutations previously discussed in this manuscript were the result of sequencing error or downstream analysis, it seems likely that is again the case here. Resequencing of this region is essential before considering *TatC* (*SecY*) as a true loss in *Ce. japonicum* and the addition of RNA sequence data would help to confirm or reject this hypothesis.

Some Genes Have Degraded into Pseudogenes

Even though secondary analysis of published sequences combined with subsequent PCR and resequencing efforts have found many of the genes that have been reported missing, this is not the case for all losses. The *rpl20* gene seems to blur the lines of deciphering when a gene is lost, and it appears to be the least conserved gene in Florideophyceae mitochondria. Interestingly, aside from its presence in red algal mitochondria, the only other lineage of eukaryotes reported to maintain *rpl20* are the jakobids (Burger and Nedelcu 2012). Retaining up to 67 genes, the most of any known mitochondria, Jakobid mitochondria are considered to most closely resemble the alpha-proteobacteria endosymbiont that became the contemporary mitochondrion (Gray et al. 2004; Burger and Nedelcu 2012; Burger et al. 2013). In the Florideophyceae, *rpl20* has been reported missing or a pseudogene in 11 species including the two new additions from this study.

Annotation of *rpl20* has been complicated because, in addition to ATG, which is the most commonly used initiation-codon for Florideophyceae mitochondrion genes, it appears that ATA may serve as an initiation-codon for *rpl20* in *Ah. plicata*, and ATT in *Rhodymenia pseudopalmata* and *Sporolithon durum* (table 4). Without these alternative initiation-codons, *rpl20* is likely a pseudogene in *Ah. plicata*, *R. pseudopalmata*, and *Sp. durum.* In addition to the aforementioned species, a conserved copy of *rpl20* using the ATG start codon was located in *Schimmelmannia schousboei* (previously not annotated) and *Schizymenia dubyi* (previously annotated as a hypothetical protein CDS).

In *Ce. japonicum, Ce. sungminbooi, C. polysiphoniae, Gelidium elegans, Gelidium vagum, H. rubra*, and *V. lanosa*, the 3′ region of *rpl20* gene remains somewhat conserved, however the 5′ end of the sequence is laden with stop codons, or appears to be missing entirely. Therefore *rpl20* is considered a pseudogene in these species. No substantial

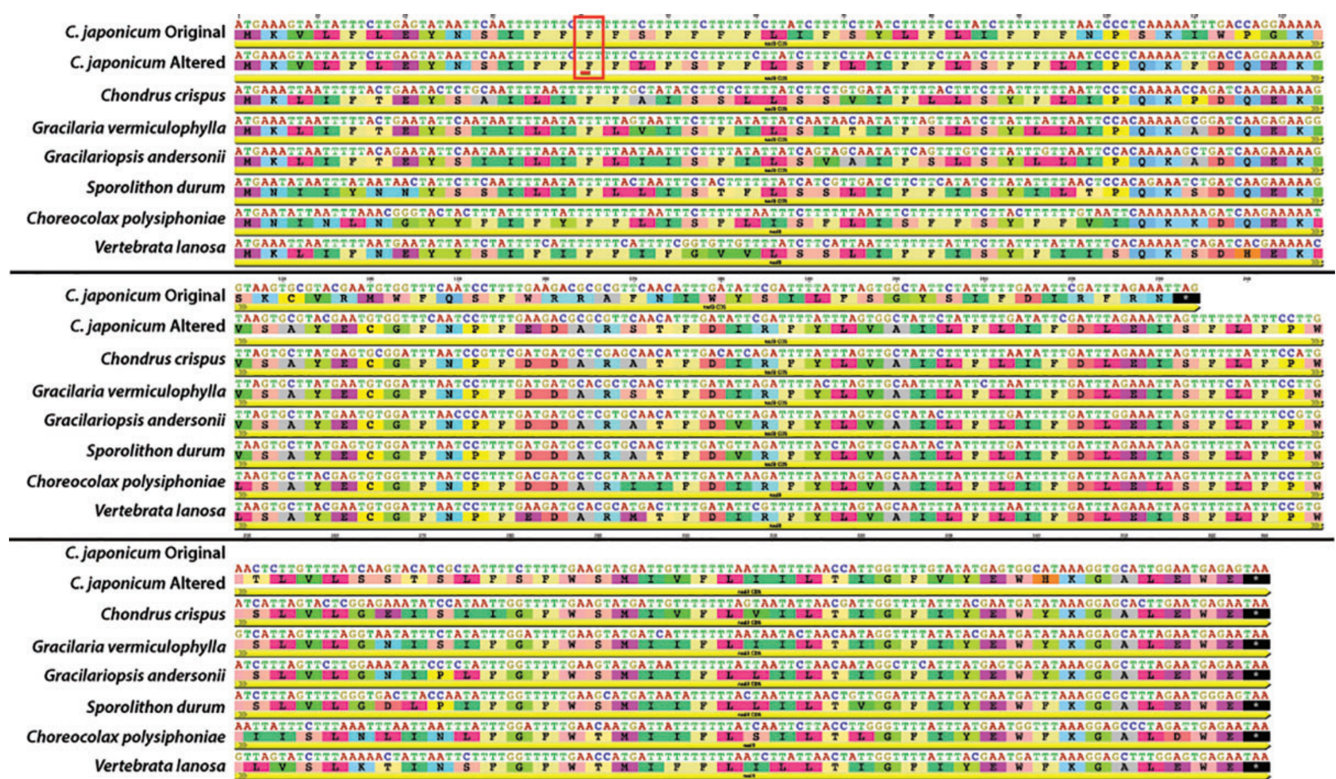

Fig. 2.—Alignment of the original *Ce. japonicum nad3* gene with the modified *Ce. japonicum nad3* ("T" deleted from base 36; red box) and copies of the *nad3* gene from *Ch. crispus, Gracilaria vermiculophylla, G. andersonii, Sp. durum, C. polysiphoniae*, and *V. lanosa*. Manual deletion of one 'T' from the string of 26 "T"s and 3 "C"s between 32 and 60 bp from the start codon restores conservation of the length and sequence of the *Ce. japonicum nad3* gene. Genes are shown with the amino acid translation below.

region in the *D. binghamiae* mitochondrial genome appears to be homologous to the *rpl20* gene. Furthermore, *rpl20* is annotated as a gene/CDS in the *Co. compressa* mitochondrial genome, however the 3′ region is slightly truncated and not highly conserved with other *rpl20* copies, suggesting that perhaps this also is a pseudogene. This wide variability in *rpl20* initiation codons and conservation cause annotation to be extremely difficult. Confirming the presence or absence of a functional *rpl20* localized in the mitochondrion is difficult and will likely require RNA sequencing and nuclear genome sequencing to identify possible cases of EGT.

The only unique Florideophyceae mitochondrion gene loss that appears to stand up to further scrutiny also encodes a ribosomal protein. Based on the published sequence of the *Sp. durum* mitochondrion, *rps11* has degraded to a pseudogene. In all other Florideophyceae, *rps11* is found adjacent to the 3′ end of *nad3*; however in *Sp. durum*, this region contains no ORFs that can be attributed to a full-length copy of *rps11*. As in other genes, a frameshift mutation appears to be initially responsible for *rps11* becoming a pseudogene. However, in all previously discussed frameshift derived pseudogenes, it was apparent that the insertion or deletion of a nucleotide or two would 'repair' the gene and result in a conserved copy that could then subsequently be confirmed by PCR. In the case of the *Sp. durum rps11*, artificially "fixing" the gene could restore a conserved 3′ end of the gene; however, a six residue gap upstream of this "fix" remained in translated alignments adding further support that *rps11* is no longer functional in *Sp. durum*. RNA and nuclear genome sequencing work remains necessary to identify whether this is a complete loss or a case of EGT from mitochondrion to the nucleus.

The Importance of Nomenclature

Identifying gene losses in Florideophyceae mitochondrial genomes has been further complicated by the use of two different names for a homologous gene. In Yang et al. (2015), the *ymf39* gene was reported as the most widely lost gene in Florideophyceae mitochondria, and was noted as being absent in six species: *Ce. japonicum*, *G. andersonii*, *H. rubra*, *Kappaphycus striatus, Sc. dubyi*, and *S. flabellata*. Furthermore, this gene is annotated only as a hypothetical protein CDS in *Gracilariopsis chorda*. Resequencing of the *G. andersonii* (KX687878) and reanalysis of the published *H. rubra* data reveals that the *ymf39* gene is present in both mitochondria. Interestingly, the other four species lacking

ymf39 are also the only Florideophyceae mitochondria with an annotated *atp4* gene, which is found between the *cox3* and *cob* genes, the same location as *ymf39* in other Florideophyceae mitochondria (Yang et al. 2015). According to Burger et al. (2003), *ymf39* encodes subunit b of mitochondrial F_0F_1-ATP synthase and should formally be designated as *atp4*.

Although it has not led to reports of gene loss, it is of note that three names have been applied to the *TatC* gene in Florideophyceae mitochondria. In *Chondrus crispus* the gene currently annotated as *ymf16* was initially described as a gene of unknown function called ORF 262 (Leblanc et al. 1995). In the publication of the *Porphyra purpurea* mitochondrial genome it was noted that *ymf16* is recognized as a homolog of *E. coli TatC* encoding a protein in the Sec-independent protein translocation pathway (Burger et al. 1999). *Coeloseira compressa, D. binghamiae*, and *K. striatus* use the name *TatC*, which is a sec-independent protein translocase protein. In all other published Florideophyceae mitochondria this gene is called *SecY*, a sec dependent protein translocase protein. When these sequences are searched against the Pfam database, all similarity hits match sec-independent protein translocase protein (*TatC*). This gene was initially incorrectly annotated as *SecY* with the publication of the second, third, and fourth florideophycean mitochondrion genomes (Hancock et al. 2010) with subsequent sequencing efforts transferring that nomenclature throughout the Florideophyceae. Furthermore, in their review of algal mitochondrial genomes, Burger and Nedelcu (2012) note that *SecY* is not found in the mitochondrial DNA of algae.

It seems reasonable for annotation efforts to rely largely upon previous publications as a reference, however in the case of *atp4*, our understanding of gene function surpassed the nomenclatural usage. The annotation of the *TatC* gene as *SecY* may have been the result of available comparative data or knowledge of mitochondrial translocase proteins at the time of the initial publication. In this effort to correct the course of mitochondrial genome annotations we support following the recommendations of Burger et al. (2003) that all *ymf39* annotations in Florideophyceae mitochondria be updated to *atp4* to reduce further confusion. Additionally, it is recommended that *SecY* and *ymf16* annotations be changed to *TatC*.

The Use of Alternative Initiation-Codons

The most widespread initiation-codon in Florideophyceae mitochondrion genes is ATG, though some exceptions have been previously proposed (table 4). For example, in the *Grateloupia angusta* mitochondria the use of ATT, TTA, or TTG as initiation-codons was reported for nine genes (Kim et al. 2014). Further examination of the published *Gra. angusta* mitochondrial genome revealed that seven of the genes reported with an alternative initiation-codon (*atp4* (as *ymf39*), *atp6, cox2, orf-Gang5, rps11, sdhB*, and *sdhC*), an ORF starting with ATG could be found within a few basepairs of the previously annotated gene, and the current *Gra. angusta atp4* (as *ymf39*) gene annotation on GenBank does use an ATG start-codon. The reasoning behind the decision to opt for an alternative codon rather than ATG at the beginning of the gene was not described in the genome announcement.

Alternative start-codons have been suggested in a few other Florideophyceae mitochondrion genes besides *Gra. angusta*. For example, in *A. taxiformis* the *atp4* gene is annotated with the initiation-codon ATT, yet 6 bp (two amino acid residues) away in the same reading frame is an ATG, which could also serve as the initiation-codon (table 4). The *Gra. angusta* and *Sp. durum* copies of *atp6* are both annotated to start with ATT codons that are 9 and 3 bp (3 and 1 amino acid residues), respectively, upstream of an ATG (table 4). A complete assessment of Florideophyceae mitochondrion genes that have been annotated using non-ATG protist mitochondrion initiation-codons and their proximity to a potential ATG start-codon is shown in table 4.

Even though some of the alternative start-codon usage is questionable, though not necessarily incorrect, there appear to be several Florideophyceae mitochondrion genes that likely are using alternative start-codons (table 4). In several of these cases the alternative hypothesis is that the genes are severely truncated and have been rendered non-functional. For example, it seems much more reasonable to believe that *S. flabellata* utilizes ATT as a start-codon for *cob* as opposed to losing the need to transcribe the first 42 amino acids of the protein. A similar situation occurs with the *sdhB* gene in *Gra. angusta*, where reliance on an ATG initiation-codon would result in the first 49 amino acids not being transcribed. Furthermore, maintaining a functional *TatC* (annotated as *SecY*) and *rpl16* in the *Gra. angusta* mitochondrion is dependent on the use of alternative initiation-codons (ATT and TTA, respectively) (table 4).

The *Ch. crispus TatC* (annotated as *ymf16*) gene is also likely reliant on an alternative start codon. Currently the *Ch. crispus TatC* gene is annotated with a GTT initiation codon. However, GTT has not been used as an initiation codon in any other Florideophyceae mitochondrion gene, nor is it one of the start codon options in translation table 4 (Protozoa Mitochondrion). Four other ORFs in the same reading frame that use either ATA or TTA as a start codon for *TatC* gene are found from 12 to 39 nucleotides downstream of the GTT codon. All *Ch. crispus* ORFs that can be reasonably attributed to *TatC* use a start-codon other than ATG, suggesting that this is another reliable instance for invoking an alternative, however the use of GTT is questionable.

Why Were Genes Reported Missing?

There are several technical and biological reasons that could explain the previous results of missing genes in Florideophyceae mitochondrial genomes. Each cell maintains

numerous mitochondria and some of these may in fact maintain the frameshift mutations that have led to gene losses being reported in published literature (Hancock et al. 2010; Yang et al. 2015). Preferential amplification of these mitochondrial genomes, or segments of the genome when using targeted PCR, would lead to the aforementioned findings even in cases where other mitochondria in the cell remain fully functional. The first four Florideophyceae mitochondrial genomes to be sequenced were completed primarily using nuclease digestions or PCR amplification and Sanger sequencing methods to assemble the genome at ~2× depth (Leblanc et al. 1995; Hancock et al. 2010). The advances in sequencing technologies and reduction in costs since the *Ch. crispus* mitochondrion was first sequenced over 20 years ago have enabled much greater sequencing depths. For example, the *V. lanosa* mitochondrial genome published here has average read coverage of 391×. This increased depth allows for the correction of "errors" either in the biology or technical aspects of sequencing by utilizing the deeper coverage of sequencing reads when forming a consensus sequence. It is noteworthy that all frameshift mutations that have been reported and led to missing genes, pseudogenes or the inversion of the *G. andersonii rps11* were found in long homopolymer regions. Assembling sequences containing long regions of low-complexity, often dominated by a single nucleotide, has been recognized as a major complication for sequencing (Kieleczawa 2006; Laehnemann et al. 2016) and was especially difficult for the 454 FLX technology used in Hancock et al. (2010) (Gilles et al. 2011; Loman et al. 2012).

RNA data can sometimes help identify problematic annotation or assembly. The apparent *atp8* and *sdhC* pseudogenes observed in their *Gr. oryzoides* DNA data were confusing to the authors as they noted that both genes were still being transcribed based on RNA sequencing efforts (Hancock et al. 2010). However, transfer to the nucleus was invoked at the time as a possible explanation. In retrospect, the RNA data was a strong indication to reexamine the data assembly. The sequencing errors reported in the first few published red algal mitochondrial genomes formed the foundation that was used as a reference for the annotation of subsequently sequenced Florideophyceae mitochondrial genomes. The apparent flexibility of mitochondrial genomes based on early sequencing efforts set a precedent for gene loss in Florideophyceae mitochondria. Building on a flawed foundation has allowed for the gene loss to be overstated without a deeper reanalysis of results. This is in no way meant as a criticism of the researchers themselves and it is plausible that results shown in data being published with current technology will be revised with future advances.

Conclusions

A detailed investigation of previously reported gene losses in Florideophyceae mitochondria reveals that losses are much less common and widespread than the published literature indicates. Prior to this study there genes had been either described as lost, or annotations were overlooked from 18 of the 30 published mitochondrial genomes (tables 1 and 3). Thoroughly examining each loss using the available published sequence data, in combination with resequencing those specimens that we could obtain material from, has positively identified 23 of the "missing" genes. Overwhelmingly, the "missing" genes or pseudogenes were the result of overlooked ORFs in the available sequence data and artificial frameshift mutations that resulted from sequencing and/or downstream assembly and analysis. In light of these findings, it is essential to thoroughly investigate results that indicate genes are degrading into pseudogenes or being lost entirely.

The *Ce. japonicum* mitochondrion was described as having lost five genes (*atp4, rpl20, sdhC, sdhD*, and *TatC*) of which three were identified here. Additionally, the translation of the existing annotation of the *nad3* gene shares no homology with any other gene, though that homology is restored through the manual deletion of a "T" in a low complexity region of that gene. Furthermore a gap in the sequence is annotated between *cob* and *nad6*. Resequencing this mitochondrion to confirm the presence or absence of the *rpl20* and *TatC*, and close the gap is essential prior to inferring biological relevance resulting from these losses. However it does seem likely that at least *rpl20* remains absent from the *Ce. japonicum* mitochondrion considering it has also been truncated in all other Ceramiales mitochondrial genomes.

It is logical that gene losses would be rare in red algal mitochondria since the core genes encoded are essential for cellular respiration and oxidative phosphorylation. The loss of genes involved in these processes would interfere with the organisms' ability to produce cellular energy and would likely be a lethal mutation. The ribosomal proteins *rps11* and *rpl20* have been lost in the mitochondria of other lineages (Burger and Nedelcu 2012) and may be examples of gene transfer from the mitochondrion to the nucleus. Additional red algal genome data will allow for the identification of nuclear-encoded, mitochondrial target proteins.

Acknowledgments

Funding for this work was provided to CL under grant no. 1257472 from the National Science Foundation. This research is based in part upon work conducted using the Rhode Island Genomics and Sequencing Center which is supported in part by the National Science Foundation (MRI grant no. DBI-0215393 and EPSCoR Grant Nos. 0554548 and EPS-1004057), the US Department of Agriculture (Grant Nos. 2002-34438-12688 and 2003-34438-13111), and the University of Rhode Island.

Literature Cited

Bayley J-P, Devilee P, Taschner PEM. 2005. The SDH mutation database: an online resource for succinate dehydrogenase sequence variants involved in pheochromocytoma, paraganglioma and mitochondrial complex II deficiency. BMC Med Genet. 6:39.

Bayley J-P, et al. 2006. Mutation analysis of SDHB and SDHC: novel germline mutations in sporadic head and neck paraganglioma and familial paraganglioma and/or pheochromocytoma. BMC Med Genet. 7:1.

Bhattacharya D, Yoon HS, Hackett JD. 2004. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. Bioessays. 26:50–60.

Blouin NA, Lane CE. 2012. Red algal parasites: models for a life history evolution that leaves photosynthesis behind again and again. Bioessays 34:226–235.

Blouin NA, Lane CE. 2015. Red algae provide fertile ground for exploring parasite evolution. Perspect Phycol. 3:11–19.

Burger G, Gray MW, Forget L, Lang BF. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biol Evol. 5:418–438.

Burger G, Lang BF, Braun HP, Marx S. 2003. The enigmatic mitochondrial ORF ymf39 codes for ATP synthase chain b. Nucleic Acids Res. 31:2353–2360.

Burger G, Nedelcu AM. 2012. Mitochondrial Genomes of Algae. In: Bock R, Knoop V, editors. Genomics of Chloroplasts and Mitochondria (Advances in photosynthesis and respiration). Dordrecht (The Netherlands): Springer, p. 127–157.

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 11:1675–1694.

Cianciola EN, Popolizio TR, Schneider CW, Lane CE. 2010. Using molecular-assisted alpha taxonomy to better understand red algal biodiversity in Bermuda. Diversity 2:946–958.

Elorza A, et al. 2004. Nuclear SDH2-1 and SDH2-2 genes, encoding the iron-sulfur subunit of mitochondrial complex II in Arabidopsis, have distinct cell-specific expression patterns and promoter activities. Plant Physiol. 136:4072–4087.

Finn RD, et al. 2010. The Pfam protein families database. Nucleic Acids Res. 38:D211–D222.

Finn RD, et al. 2015. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44:D279–D285.

Gilles A, et al. 2011. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC Genomics 12:245.

Gould SB, Waller RR, McFadden GI. 2008. Plastid evolution. Annu Rev Plant Biol. 59:491–517.

Gray MW. 2012. Mitochondrial Evolution. Cold Spring Harb Perspect Biol. 4:a011403.

Gray MW, Burger G, Lang BF. 1999. Mitochondrial evolution. Science 283:1476–1482.

Gray MW, Lang BF, Burger G. 2004. Mitochondria of protists. Annu Rev Genet. 38:477–524.

Guiry MD, Guiry GM. 2016. Algaebase. Ireland, Galway: World-wide Electron. Publ. Natl. Univ. Avialable from: http//www.algaebase.org. searched on 21 July 2016.

Hancock L, Goff LJ, Lane CE. 2010. Red algae lose key mitochondrial genes in response to becoming parasitic. Genome Biol Evol. 2:897–910.

Hughey JR, Boo GH. 2016. Genomic and phylogenetic analysis of *Ceramium cimbricum* (Ceramiales, Rhodophyta) from the Atlantic and Pacific Oceans supports the naming of a new invasive Pacific entity *Ceramium sungminbooi* sp. nov. Bot Mar. 59:211–222.

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:e59001.

Karnkowska A, et al. 2016. A eukaryote without a mitochondrial organelle. Curr Biol. 26:1274–1284.

Kearse M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Keeling PJ. 2004. Diversity and evolutionary history of plastids and their hosts. Am J Bot. 91:1481–1493.

Keeling PJ. 2010. The endosymbiotic origin, diversification and fate of plastids. Philos Trans R Soc Lond B Biol Sci. 365:729–748.

Kieleczawa J. 2006. Fundamentals of sequencing of difficult templates-An overview. J Biomol Tech. 17:207–217.

Kim SY, Yang EC, Boo SM, Yoon HS. 2014. Complete mitochondrial genome of the marine red alga *Grateloupia angusta* (Halymeniales). Mitochondrial DNA. 25:269–270.

Koonin EV. 2010. The origin and early evolution of eukaryotes in the light of phylogenomics. Genome Biol. 11:209.

Ku C, Nelson-Sathi S, Roettger M, Garg S, et al. 2015. Endosymbiotic gene transfer from prokaryotic pangenomes: Inherited chimerism in eukaryotes. Proc Natl Acad Sci. USA 112:10139–10146.

Ku C, Nelson-Sathi S, Roettger M, Sousa FL, et al. 2015. Endosymbiotic origin and differential loss of eukaryotic genes. Nature 524:427–432.

Laehnemann D, Borkhardt A, McHardy AC. 2016. Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction. Brief Bioinform. 17:154–179.

Lam DW, Verbruggen H, Saunders GW, Vis ML. 2016. Multigene phylogeny of the red algal subclass Nemaliophycidae. Mol Phylogenet Evol. 94:730–736.

Lane CE, Archibald JM. 2008. The eukaryotic tree of life: endosymbiosis takes its TOL. Trends Ecol Evol. 23:268–275.

Lang BF, Gray MW, Burger G. 1999. Mitochondrial genome evolution and the origin of Eukaryotes. Annu Rev Genet. 33:351–397.

Leblanc C, Kloareg B, Loiseaux-de Goër S, Boyen C. 1995. DNA sequence, structure, and phylogenetic relationship of the mitochondrial small-subunit rRNA from the red alga *Chondrus crispus* (Gigartinales, Rhodophytes). J Mol Evol. 41:196–202.

Le Gall L, Saunders GW. 2010. DNA barcoding is a powerful tool to uncover algal diversity: a case study of the Phyllophoraceae (Gigartinales, Rhodophyta) in the Canadian flora. J Phycol. 46:374–389.

Loman NJ, et al. 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 30:434–439.

Martin W, et al. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature 393:162–165.

Martin WF, Garg S, Zimorski V. 2015. Endosymbiotic theories for eukaryote origin. Philos Trans R Soc B. 370:20140330.

Qiu H, et al. 2013. Assessing the bacterial contribution to the plastid proteome. Trends Plant Sci. 18:680–687.

Qiu H, Price DC, Yang EC, Yoon HS, Bhattacharya D. 2015. Evidence of ancient genome reduction in red algae (Rhodophyta). J Phycol. 51:624–636.

Salomaki ED, Lane CE. 2014. Are all red algal parasites cut from the same cloth?. Acta Soc Bot Pol. 83:369–375.

Saunders GW. 1993. Gel purification of red algal genomic DNA: an inexpensive and rapid method for the isolation of polymerase chain reaction-friendly DNA. J Phycol. 29:251–254.

Saunders GW, McDonald B. 2010. DNA barcoding reveals multiple overlooked Australian species of the red algal order Rhodymeniales (Florideophyceae), with resurrection of *Halopeltis* J. Agardh and description of *Pseudohalopeltis* gen. nov. Bot Botanique. 88:639–667.

Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33:686–689.

Smith DR, Keeling PJ. 2015. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. Proc Natl Acad Sci. USA 112:10177–10184.

Stiller JW. 2007. Plastid endosymbiosis, genome evolution and the origin of green plants. Trends Plant Sci. 12:391–396.
Stiller JW, et al. 2014. The evolution of photosynthesis in chromist algae through serial endosymbioses. Nat Commun. 5:1–7.
Tamayo DA, Hughey JR. 2016. Organellar genome analysis of the marine red alga *Dasya binghamiae* (Dasyaceae, Rhodophyta) reveals an uncharacteristic florideophyte mitogenome structure. Mitochondrial DNA Part B. 1:510–511.
Tanifuji G, Archibald JM, Hashimoto T. 2016. Comparative genomics of mitochondria in chlorarachniophyte algae: endosymbiotic gene transfer and organellar genome dynamics. Sci Rep. 6:21016.
Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 5:123–135.
Tucker RP. 2013. Horizontal gene transfer in choanoflagellates. J Exp Zool B Mol Dev Evol. 320:1–9.
Verbruggen H, et al. 2010. Data mining approach identifies research priorities and data requirements for resolving the red algal tree of life. BMC Evol Biol. 10:16.
Yang E, et al. 2015. Highly conserved mitochondrial genomes among multicellular red algae of the Florideophyceae. Genome. Biol Evol. 7:2394–2406.
Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol. 21:809–818.
Zimorski V, Ku C, Martin WF, Gould SB. 2014. Endosymbiotic theory for organelle origins. Curr Opin Microbiol. 22:38–48.

Associate editor: Martin Embley