

The inflated mitochondrial genomes of siphonous green algae reflect processes driving expansion of noncoding DNA and proliferation of introns

Sonja I. Repetti¹, Christopher J. Jackson¹, Louise M. Judd², Ryan R. Wick², Kathryn E. Holt² and Heroen Verbruggen¹

¹ School of BioSciences, University of Melbourne, Melbourne, VIC, Australia

² Department of Infectious Diseases, Monash University, Melbourne, VIC, Australia

ABSTRACT

Within the siphonous green algal order Bryopsidales, the size and gene arrangement of chloroplast genomes has been examined extensively, while mitochondrial genomes have been mostly overlooked. The recently published mitochondrial genome of *Caulerpa lentillifera* is large with expanded noncoding DNA, but it remains unclear if this is characteristic of the entire order. Our study aims to evaluate the evolutionary forces shaping organelle genome dynamics in the Bryopsidales based on the *C. lentillifera* and *Ostreobium quekettii* mitochondrial genomes. In this study, the mitochondrial genome of *O. quekettii* was characterised using a combination of long and short read sequencing, and bioinformatic tools for annotation and sequence analyses. We compared the mitochondrial and chloroplast genomes of *O. quekettii* and *C. lentillifera* to examine hypotheses related to genome evolution. The *O. quekettii* mitochondrial genome is the largest green algal mitochondrial genome sequenced (241,739 bp), considerably larger than its chloroplast genome. As with the mtDNA of *C. lentillifera*, most of this excess size is from the expansion of intergenic DNA and proliferation of introns. Inflated mitochondrial genomes in the Bryopsidales suggest effective population size, recombination and/or mutation rate, influenced by nuclear-encoded proteins, differ between the genomes of mitochondria and chloroplasts, reducing the strength of selection to influence evolution of their mitochondrial genomes.

Submitted 2 October 2019
Accepted 22 November 2019
Published 3 January 2020

Corresponding author
Sonja I. Repetti,
sonjarepetti@live.com.au

Academic editor
Jürg Bähler

Additional Information and
Declarations can be found on
page 16

DOI 10.7717/peerj.8273

© Copyright
2020 Repetti et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Bioinformatics, Evolutionary Studies, Genomics, Marine Biology, Molecular Biology
Keywords *Ostreobium*, Bryopsidales, Siphonous, Chlorophyta, Genome evolution, Mitochondria, Natural selection, Genetic drift, Organelles

INTRODUCTION

Although they share many unifying features, the green algae (Chlorophyta) are a diverse group with genomes that vary considerably in structure and gene content (*Yurina & Odintsova, 2016*). Most work on Chlorophyta genomes to date has focused on organellar genomes, with in excess of 150 chloroplast and 70 mitochondrial genomes published on GenBank. These organelle genome data have been used for phylogenetic analyses to resolve lineage relationships within the Chlorophyta (*Fučíková et al., 2014; Cremen et al., 2018*) and

to investigate organellar genome structural adaptation and unique features. As with many plastid-bearing taxa, more effort has been made to sequence and characterise chloroplast genomes rather than mitochondrial, as plastid genomes tend to be favoured for their use in phylogenetic studies due to high coding content, relatively conserved genome and simple inheritance patterns (Smith & Keeling, 2015; Fang et al., 2017). There are some notable exceptions in the Chlorophyta such as the fragmented hairpin plasmids seen in Cladophorales plastids (Del Cortona et al., 2017) and the large and highly repetitive *Acetabularia acetabulum* (Dasycladales) plastid genome (De Vries et al., 2013). Despite this, mitochondrial genomes tend to show greater abnormalities in structure than plastid genomes (Burger, Gray & Lang, 2003), and some of the most unusual genomic features have been observed in eukaryotic mitochondrial genomes (Smith & Keeling, 2015).

Comparative studies in green algal chloroplast and mitochondrial genomes have found that many instances of genome size variation are not due to gene gain or loss, but rather reduction or expansion of noncoding DNA including introns (Burger, Gray & Lang, 2003; Smith et al., 2013; Marcelino et al., 2016). In the Chlamydomonadales, Smith et al. (2013) observed an increase in chloroplast DNA from approximately 60% noncoding DNA in smaller-sized unicellular lineages to greater than 80% in multicellular *Volvox carteri*, while the number of genes differs by only two between the smallest and largest genomes. A similar pattern was observed for their mitochondrial DNA (Smith et al., 2013). According to Smith & Lee (2010), the large noncoding DNA content in *V. carteri* genomes could support the mutational-hazard hypothesis (MHH). The MHH proposes that excess DNA is more likely to accumulate in genomes with a low mutation rate and small effective population size (N_e) (Lynch, 2006; Smith, 2016) due to noncoding DNA imposing a fitness burden by increasing the potential for harmful mutations. Such excess DNA can be eradicated more effectively when there is a larger effective population size for natural selection to act upon through purifying selection (Lynch & Walsh, 2007). The MHH might explain the streamlining of organelle genomes within various lineages including prasinophytes, red algae, and some fungi, which have high estimated mutation rates (Smith, 2016). A large reduction in N_e is expected for multicellular organisms compared with unicellular, therefore purifying selection should be less efficient meaning there is less restriction on the accumulation of noncoding DNA. This may explain the difference in noncoding content between unicellular members of the Chlamydomonadales, such as *Chlamydomonas*, and colonial *V. carteri* (Smith et al., 2010). Various studies have supported and challenged the MHH (Smith, 2016), but on the whole emphasise the importance of mutation and genetic drift in shaping organelle genomes (Lynch, 2006; Smith, 2016).

It has also been suggested that selection can influence the mutation rate of genomes. The drift-barrier hypothesis (Lynch et al., 2016) proposes that selection acts to reduce the mutation rate with an overall limit set by genetic drift. In genomes with very high mutation rates, 'antimutators' such as DNA repair proteins are advantageous and decrease the mutation rate until the strength of selection is matched by that of genetic drift and mutation bias (Lynch et al., 2016). In contrast, in genomes with low mutation rates, antimutators will not be advantageous enough to be selected for, whilst having mild

mutators will not be sufficiently disadvantageous to be selected against. Therefore, mutation rate will increase until selection is strong enough to prevent the genome evolving a higher mutation rate (Lynch *et al.*, 2016). Krasovec *et al.* (2017) combined a review of literature with mutation rate estimates for four prasinophytes and found support for this theory, with mutation rate tending to decrease as N_e , therefore strength of selection, increases.

The effective population size of genomes influences the efficacy of selection relative to genetic drift (Ness *et al.*, 2016). Within genomes, loci linked to sites under selection will exhibit locally reduced N_e due to hitchhiking, where mutations that are not necessarily optimal are fixed because they are located close to another gene undergoing selection (Smith & Haigh, 1974), background selection, when selection acts against linked deleterious alleles that are retained in a population by mutation (Charlesworth, 1994), and Hill–Robertson (HR) interference (Platt, Weber & Liberles, 2018 and references therein). Under the HR effect (Hill & Robertson, 1966), weak selection acting on linked sites reduces the overall effectiveness of selection due to linkage disequilibrium (Comeron, Williford & Kliman, 2008). Recombination breaks up such linkage and prevents local reduction in N_e due to selection acting upon linked sites (Cutter & Payseur, 2013; Platt, Weber & Liberles, 2018). It has been observed that genes located in regions of low recombination have a greater number of introns, and these introns have greater lengths (Comeron & Kreitman, 2000). It has been proposed that introns may be advantageous in reducing intragenic HR effect and are hence retained, as they might increase rates of recombination that lower intragenic HR effects (Comeron, Williford & Kliman, 2008). Comeron & Kreitman (2000) suggest that longer introns are advantageous in regions of low recombination, as they reduce the HR effect relative to a shorter intron by increasing the recombination rate between mutations in different exons, and that they will be maintained due to hitchhiking effects between favourable mutations and the longest intron variant.

The class Ulvophyceae within the Chlorophyta show high morphological diversity, including a range of cell configurations (Cocquyt *et al.*, 2010; Fang *et al.*, 2017), and are thus a useful model for examining the evolution of genome features under differing evolutionary constraints. The order Bryopsidales are a lineage of siphonous seaweeds with thali comprised of a single giant tubular cell containing cytoplasm with a large number of nuclei and other organelles free to move around the entire plant (Vroom & Smith, 2001, 2003; Verbruggen *et al.*, 2009; Mine, Sekida & Okuda, 2015). Within the Bryopsidales, the size and gene arrangement of chloroplast genomes has been found to vary considerably while gene content was much less variable (Cremen *et al.*, 2018). While appreciable short-read sequencing data have been generated for the Bryopsidales, mitochondrial genomes have not been found to assemble well from the sequencing libraries used to assemble chloroplast genomes, and the first bryopsidalean mitochondrial genome was published only very recently for the sea grape *Caulerpa lentillifera* (Zheng *et al.*, 2018). This 209,034 bp circular-mapping genome is the largest Chlorophyta mitochondrial genome published to date, an order of magnitude larger than the average for Chlorophyta mitochondrial genomes and rich in intergenic sequences and introns, the causes of which have not been examined.

Unlike the genome expansion observed in the Chlamydomonadales, which appears to occur in both organellar compartments (Smith *et al.*, 2013; Zhang *et al.*, 2019), *C. lentillifera*'s mitochondrial genome is much larger than its chloroplast genome. This is unusual for green algae where mitochondrial genomes are generally smaller than chloroplast genomes (Leliaert *et al.*, 2012). Furthermore, mitochondrial genome expansion is less frequent in green algae compared with plants (Mower, Sloan & Alverson, 2012). While Zheng *et al.* (2018) did not propose an evolutionary explanation for these findings, the inflated noncoding regions described are consistent with previous studies in plants showing that the variation in the sizes of their mitochondrial genomes is predominantly explained by changes in noncoding DNA content including repeats, introns, intergenic DNA and DNA of foreign origin (Yurina & Odintsova, 2016). Therefore, the mitochondrial genome of *C. lentillifera* provides an interesting case in the Chlorophyta to examine the evolution of organellar genomes and, in particular, the processes driving genome expansion. However, it remains undetermined whether a large mitochondrial genome is a feature only of *Caulerpa*, or if it is in fact characteristic of the order Bryopsidales.

In this paper, we focus on another member of the Bryopsidales that is currently of considerable ecological interest: the genus *Ostreobium*, an endolithic limestone-boring alga (Verbruggen & Tribollet, 2011). *Ostreobium* is found in a diverse range of calcium carbonate environments around the world, is abundant in the skeleton of many coral species, and it is one of the most common genera of endolithic autotrophs in coral reefs (Tribollet, 2008). Endolithic algae like *Ostreobium* have a number of observed and predicted roles in the coral skeleton (Ricci *et al.*, 2019). While their boring can destabilise coral skeletons and impact on reef structure (Schlichter, Kampmann & Conrady, 1997), they may have a mutualistic relationship with corals, providing them with metabolites that could enable them to survive bleaching events (Schlichter, Kampmann & Conrady, 1997; Tribollet, 2008). Its endolithic lifestyle means that *Ostreobium* inhabits low light habitats with limited available photosynthetically active radiation (Wilhelm & Jakob, 2006; Magnusson, Fine & Kühl, 2007) and it has been shown to have far red-shifted absorption spectra, allowing it to use the long wavelengths of light available in the coral skeleton (Wilhelm & Jakob, 2006; Magnusson, Fine & Kühl, 2007; Tribollet, 2008).

Ostreobium species have consistently small chloroplast genomes relative to the median for Bryopsidales of 105 kb (Cremen *et al.*, 2018), with the chloroplast genome of *Ostreobium* strain HV05042 the most compact found so far (80,584 bp) in the Ulvophyceae (Marcelino *et al.*, 2016; Verbruggen *et al.*, 2017). Marcelino *et al.* (2016) identified only three introns in the *Ostreobium quekettii* chloroplast genome, and there was an overall reduction in intergenic regions that resulted in its reduced size. They hypothesised that this might reflect evolution in response to resource constraints in its low light environment (Marcelino *et al.*, 2016), but did not examine the mitochondrial genome for evidence of similar selection pressures.

The goal of this study is to evaluate the characteristics of mitochondrial genomes in the Bryopsidales, as well as to determine the evolutionary forces that have shaped organelle genome dynamics in the Bryopsidales. Our approach consists of the sequencing, assembly

on the results of BLASTx searches of sections of this candidate contig against the NCBI nr and nt databases.

This contig was used as the query in a BLASTn search against the *O. quekettii* Spades short-read assembly, within Geneious with default settings. Top hits were aligned with the long-read contig within Geneious. Scaffolds from the short-read assembly were used as a reference to manually correct the long-read contig within Geneious. In order to verify that the genome was circular-mapping, we searched for a short-read scaffold spanning across both ends of the long-read contig when aligned.

This manual result was compared with the mitochondrial contig identified, through blast searches against the NCBI nr and nt databases, in the MaSuRCA hybrid assembly by alignment in Geneious.

Genome annotation

The genome was initially annotated using MFannot ([Beck & Lang, 2010](#)), and DOGMA ([Wyman, Jansen & Boore, 2004](#)) with very relaxed settings (protein identity cut off 25%, RNA identity cut off 30%). Annotations of predicted protein coding genes were confirmed through extraction of ORFs and BLAST searches of these against the NCBI nr and nt databases, as well as alignment with transcripts that were recovered as hits from BLASTn searches against the *O. quekettii* transcriptome within Geneious, using default settings.

tRNAs were identified using tRNAscan-SE ([Lowe & Chan, 2016](#)), tRNAfinder ([Kinouchi & Kurokawa, 2006](#)), ARAGORN ([Laslett & Canback, 2004](#)) and tRNADB-CE's BLASTN/Pattern Search ([Abe et al., 2010](#)). rRNAs were identified with RNAmmer ([Lagesen et al., 2007](#)) and RNAweasel ([Lang, Laforest & Burger, 2007](#)).

Repeats were identified using the tandem repeats database ([Gelfand, Rodriguez & Benson, 2007](#)), the RepeatFinder package in Geneious with a minimum repeat length of 50 bp ([Pombert et al., 2004](#)) and REPuter ([Kurtz et al., 2001](#)) with minimal repeat size setting of 12 bp ([Smith & Lee, 2009](#)). Forward, reverse, complement, and reverse complement repeats were all considered under REPuter.

A map of the genome was created with Circos ([Krzywinski et al., 2009](#)) and manually annotated in Inkscape 0.92 (www.inkscape.org). The sequence for the *O. quekettii* mitochondrial genome is available on GenBank (Accession number [MN514984](#)).

Open reading frames

Open reading frames were predicted using ORF finder in Geneious with a minimum length setting of 300 bp. These were used as queries in BLASTx searches against the NCBI nr and nt databases (e value = e^{-1}), and the translated ORFs were used as queries in a batch sequence search against the Pfam database ([Finn et al., 2017](#)). Only ORFs that had BLAST results and identified Pfam domains were retained in the final genome annotation. In order to determine relationships between ORFs indicative of common origin, the annotated ORFs along with those from the chloroplast genome of *O. quekettii* ([Marcelino et al., 2016](#)) and the mitochondrial genome of *C. lentillifera* ([Zheng et al., 2018](#)) were clustered based on all-against-all BLAST+ similarities using CLANs ([Frickey & Lupas, 2004](#)). This was performed through the MPI Bioinformatics toolkit ([Zimmermann et al., 2018](#)),

with the BLOSUM62 scoring matrix and extraction of BLAST HSPs up to e -values of $1e-4$. The output from CLANs was annotated in Inkscape.

Introns

Intron boundaries were inferred by aligning predicted genes with their corresponding transcripts and intron-lacking homologues from other green algae. Intron class was predicted using RNAweasel, and Rfam sequence search ([Griffiths-Jones et al., 2003](#)).

In order to identify potential common origins of introns in the *O. quekettii* and *C. lentillifera* mitochondrial genomes as well as the *O. quekettii* plastid genome, we used a distance-based clustering technique. In contrast to the analysis of ORFs above, this analysis used DNA sequences and used introns that were not interrupted by ORFs, in the hope that this would represent a subset of the data that could be more easily aligned if they contained conserved regions. From the intronic DNA sequences, a distance matrix was constructed by comparing the introns from these genomes using Clustal Omega ([Sievers et al., 2011](#)) with the ‘—distmat-out’ and ‘—full’ flags. This distance matrix was used as the input to construct a neighbour joining (NJ) tree using Neighbor within the PHYLIP package ([Felsenstein, 2005](#)). We also constructed a distance matrix using a subset of these introns from only the *O. quekettii* chloroplast and mitochondrial genomes and this was used to construct a NJ tree. This NJ tree was further visualised and annotated in MEGA ([Kumar et al., 2018](#)). Clusters of introns identified in the NJ tree were aligned in Geneious using MAFFT to better evaluate similarity. Insertion sites of introns were also compared between the *O. quekettii* and *C. lentillifera* genomes by aligning gene sequences in Geneious using MAFFT.

Rates of evolution

To obtain an estimate of the relative rates of evolution in the *O. quekettii* mitochondrial and chloroplast genomes, we aligned all the protein coding and rRNA genes common between the *O. quekettii* and *C. lentillifera* mitochondrial and chloroplast genomes in Geneious using the default aligner. We generated estimates of base substitutions per site between sequences using the Jukes–Cantor model ([Jukes & Cantor, 1969](#)) in MEGA. We also attempted to obtain d_N/d_S ratios for gene alignments using PAL2NAL ([Suyama, Torrents & Bork, 2006](#)).

Mitochondrion-targeted nuclear genes

In the context of evaluating hypotheses around mitochondrial DNA repair machinery (see below), we searched for homologues of nuclear-encoded DNA repair proteins that are targeted to the mitochondrion in higher plants: MSH1, RECA proteins and OSB1. *Arabidopsis* sequences were used as queries in BLAST searches with default settings against the nuclear contigs in the *O. quekettii* hybrid assembly, as well as the *O. quekettii* transcriptome in Geneious. We searched for putative targeting signals to the mitochondrion with DeepLoc-1.0 using default settings ([Almagro Armenteros et al., 2017](#)).

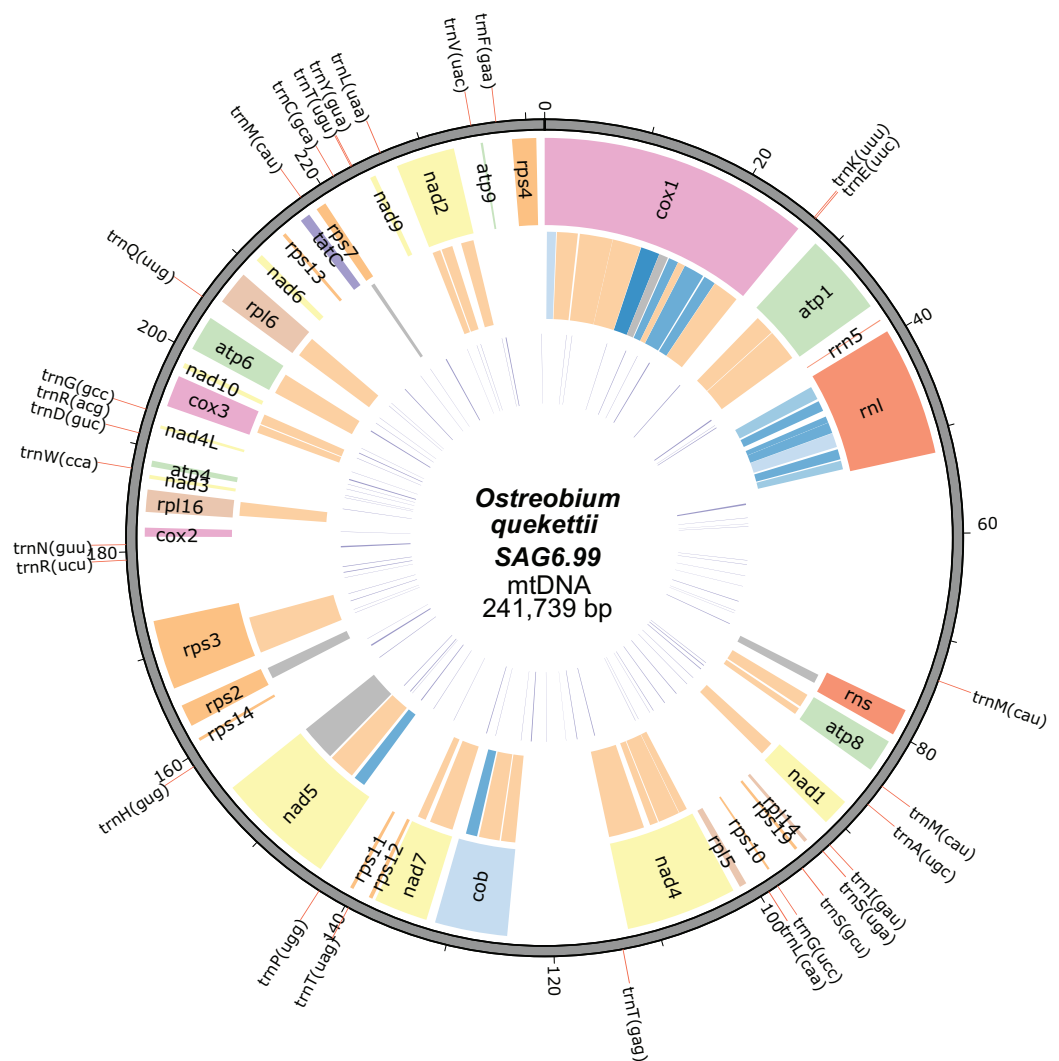


Figure 1 Mitochondrial genome map of *Ostreobium quekettii* SAG6.99. The position of tRNAs are shown on the outer track (red lines). The first inner circle represents the position, size and the names of the protein-coding and rRNA genes. The introns are shown in the second inner circle and are colour coded according to intron types/subtypes: group I derived (very light blue), group IA (light blue), group IB (blue), group ID (dark blue), group II (orange), unknown (grey). The third inner circle represents the position and length (line thickness) of repeats. [Full-size !\[\]\(ba1b80118482ccef74a5d718ca4d7242_img.jpg\) DOI: 10.7717/peerj.8273/fig-1](https://doi.org/10.7717/peerj.8273/fig-1)

RESULTS

Ostreobium mitochondrial genome

The mitochondrial genome was captured by a single contig in the Canu assembly, but six separate contigs in the Spades assembly. The genome was captured by a single contig in the hybrid genome assembly that showed 99.97% pairwise identity with our manually corrected version of the genome. The combination of long- and short-read data improved this assembly and allowed us to capture the entire genome with a high level of accuracy.

The mitochondrial genome of *O. quekettii* SAG6.99 assembled into a 241,739 bp circular-mapping molecule (Fig. 1). All 64 codons are used (Table S1) and the 28 tRNAs

Table 1 Summary of coding and noncoding content of the mtDNA of *Ostreobium quekettii* SAG6.99.

	Percent of total noncoding DNA	Percent of overall genome
Coding (rRNA, tRNA, ORFs, protein coding genes)		25%
Repeats	7%	5%
Intergenic DNA	61%	46%
Introns (including ORFs)		39%
Introns (excluding ORFs)	39%	29%
Total noncoding DNA (excluding ORFs)		75%
Total intronic and intergenic DNA (including ORFs)		85%

Table 2 Genes, introns and open reading frames present in mtDNA of *Ostreobium quekettii* SAG6.99. For further information on introns and ORFs, see [Supplemental Material](#).

	Number in genome
Protein coding genes	34
rRNA	3
tRNA	28
Genes containing introns	18
Introns	47
Group I	14
Group II	28
Unclear	5
Introns containing ORFs	18
ORFs	20
Intronic	20
Intergenic	0

encoded by the *O. quekettii* mtDNA (Table S2) appear to be sufficient to recognise all of these codons assuming the standard genetic code and maximum use of wobbling and superwobbling (Alkatib *et al.*, 2012). Only 25% of the genome is coding DNA (Table 1). The genome encodes 3 rRNAs and 28 tRNAs (Tables 2 and 3), resembling other green algal mitochondrial genomes (Table S3). It also encodes 34 named protein-coding genes commonly found in green algae (Table 3; Table S3). This does include some genes that are less common in Chlorophyta mitochondrial genomes: *nad10*, which is absent from the mtDNA of sequenced land plants and many green algae (Mower, Sloan & Alverson, 2012), and *tatC*, a gene encoding a component of the inner membrane TAT translocase responsible for transporting folded proteins across the membrane in bacteria but whose function in mitochondria remains unclear (Carrie, Weißenberger & Soll, 2016; Petru *et al.*, 2018).

The overall GC content of the genome is 48.3% (Table 1), which is higher than the average for eukaryotic mitochondrial genomes (38%) (Smith & Lee, 2008) but similar to that of *C. lentillifera* (50.9%) (Zheng *et al.*, 2018) and not extreme for green algae

Table 3 Protein coding and ribosomal genes present in the mtDNA of *Ostreobium quekettii* SAG6.99.

Protein genes	
Complex I (<i>nad</i>)	<i>nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9, nad10</i>
Complex III (<i>cob</i>)	<i>cob</i>
Complex IV (<i>cox</i>)	<i>cox1, cox2, cox3</i>
Complex V (<i>atp</i>)	<i>atp1, atp4, atp6, atp8, atp9</i>
SSU ribosomal proteins (<i>rps</i>)	<i>rps2, rps3, rps4, rps7, rps10, rps11, rps12, rps13, rps14, rps19</i>
LSU ribosomal proteins (<i>rpl</i>)	<i>rpl5, rpl6, rpl14, rpl16</i>
Ribosomal RNAs	<i>rrn5, rrnS, rrnL</i>
Putative Protein Transporter	<i>tatC</i>

(*Del Vasto et al., 2015*). It does however vastly exceed the GC content of the *O. quekettii* chloroplast genome (31.9%) (*Marcelino et al., 2016*).

The *O. quekettii* mitochondrial genome contains 373 repeats, found in both introns and intergenic DNA, that represent 5% of the total genome (*Fig. 1; Table 1*), with a minimum length of 31 and maximum of 299 bp (mean 107.3 ± 61.7 SD). This is similar to the mitochondrial genome of *C. lentillifera*, where we estimated 349 repeats represent approximately 4% of the genome, with a minimum length of 25 bp and maximum of 163 bp (mean 68.3 ± 30.1 SD).

Introns

Most of the genome's inflated size (75%) is from expanded intergenic and intronic regions rather than extra coding material. Eighteen of the 34 named protein-coding genes contain one or multiple intron(s) (*Table 2; Table S4*). Introns include both group I and group II introns, as well as five whose class could not be confidently determined (*Table 2; Table S4*). The *Ostreobium* mitochondrial genome contains 47 introns, compared with only 29 in the mtDNA of *C. lentillifera* (*Zheng et al., 2018*). Group II introns, which appear to have been abundant early in the evolution of Bryopsidales plastid genomes (*Cremen et al., 2018*) and are the dominant type in plant mtDNA but less common elsewhere (*Lang, Laforest & Burger, 2007*), are the dominant type in the *O. quekettii* mtDNA. This contrasts with the *C. lentillifera* mitochondrial genome, which also contains both types of introns but with group I introns predominating (*Zheng et al., 2018*).

Open reading frames-lacking introns from the *O. quekettii* and *C. lentillifera* mitochondrial genomes mostly form separate clusters in NJ trees of pairwise DNA distances (*Fig. S1*). Alignments of the few introns that do cluster together did not show convincing homology between *C. lentillifera* and *O. quekettii* mitochondrial genomes (alignments not shown), nor between introns in the *O. quekettii* mitochondrial and plastid genomes (*Fig. S2*). Furthermore, comparing insertion sites for introns between *O. quekettii* and *C. lentillifera* only revealed five sites where intron insertion sites of the same type of intron were within 3 bp of each other, and only a single site where they matched exactly (*Table S4*). The NJ tree did identify groups of similar group II introns within the *O. quekettii*

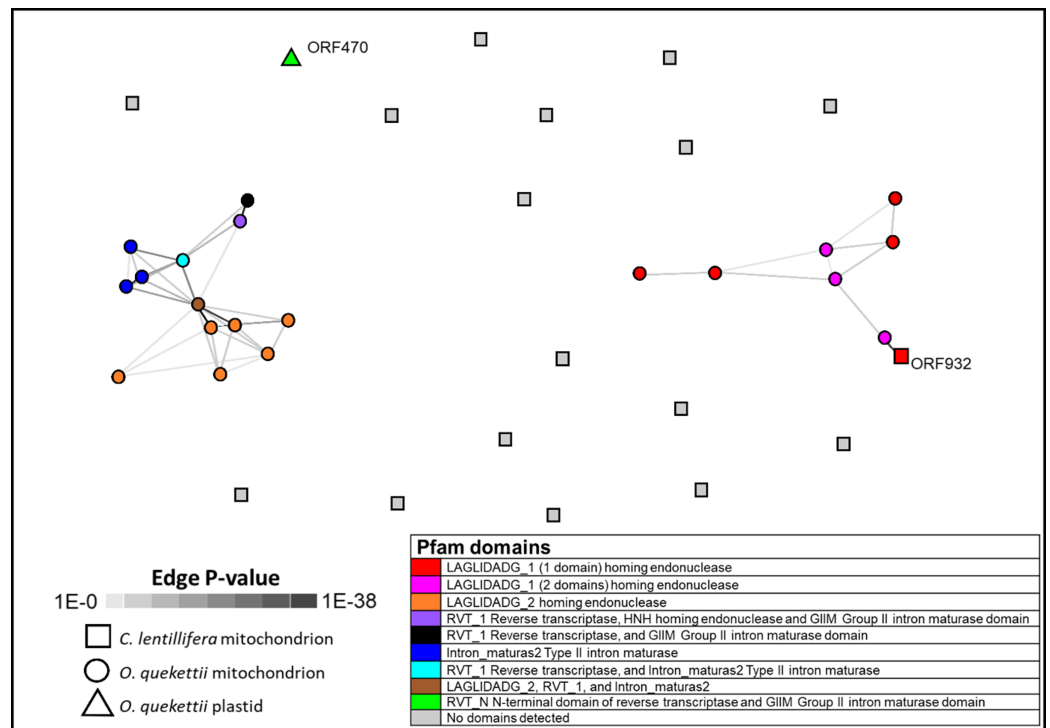


Figure 2 Similarity network generated from all against all BLAST+ similarities of ORFs encoded in the *Caulerpa lentillifera* mitochondrion, and *Ostreobium quekettii* mitochondrion and chloroplast. Each node represents an ORF, and each edge (line) represents a significant HSP (high scoring segment pair), shaded according to p value. Generated using CLANS through the MPI Bioinformatics Toolkit (Scoring Matrix BLOSSUM62, extracting BLAST HSPs up to e -values of $1e-4$, using p -values better than 1.0). [Full-size !\[\]\(ba1b80118482ccef74a5d718ca4d7242_img.jpg\) DOI: 10.7717/peerj.8273/fig-2](https://doi.org/10.7717/peerj.8273/fig-2)

mitochondrial genome, and these sequences did show convincing similarity at the DNA level in alignments, suggesting that these groups of introns proliferated within the *Ostreobium* lineage (Fig. S2).

Open reading frames with identifiable Pfam domains were only found in introns, not intergenic DNA (Table 2). Eighteen of the introns in the *O. quekettii* mitochondrial genome contain one or more ORFs. In a similarity network generated from all against all BLAST+ similarities, these ORFs appear to cluster together based on shared Pfam domains, which predict similar functions for these ORFs (Fig. 2). ORFs show homology to intron-encoded proteins that act as maturases and homing endonucleases, which enable splicing and promote intron mobility (Lambowitz & Belfort, 1993; Lambowitz & Zimmerly, 2011; Hausner, 2012). They contain a variety of domains, but mostly two types of domains with the amino-acid motif LAGLIDADG, which are common in both group I and group II introns (Hausner, 2012). Three ORFs contain double LAGLIDADG_1 domains, which in other lineages are in intron-encoded proteins with maturase activity that enable intron splicing (Lambowitz & Belfort, 1993). Four ORFs contain an RVT domain; Cremen et al. (2018) found evidence for the mobility of group II intron-encoded ORFs containing an RVT domain within Bryopsidales chloroplast genomes.

Only one ORF identified by [Zheng et al. \(2018\)](#) in *C. lentillifera* had a positive hit to the Pfam database: ORF932 in *cox1* had a putative LAGLIDADG_1 domain. This sequence clustered with *O. quekettii* ORF sequences containing the same domain ([Fig. 2](#)).

None of the *O. quekettii* mitochondrial ORFs show similarity to the single ORF460 identified in the *O. quekettii* chloroplast, which has a putative intron splicing function ([Cremen et al., 2018](#)). Along with the results for the introns, this suggests there is little evidence for the transfer of ORFs or introns between organellar genomes in *O. quekettii*.

Testing evolutionary theories: rates of evolution and mitochondrion-targeted proteins

The inflated mitochondrial genome of *O. quekettii* prompted several hypotheses (see below) that could be examined using molecular data.

In order to gain insight into whether there are differential mutation rates between the plastid and mitochondrial genomes, that might help explain their differing sizes in both *O. quekettii* and *C. lentillifera*, we aligned genes from the two organellar genomes between the two species. Estimates of base substitutions per site using a simple Jukes–Cantor model showed a slightly higher ($t = 2.94$, $p < 0.01$) number of average base substitutions per site for the mitochondrial genomes (mean 0.456 ± 0.207 SE) compared with chloroplast genomes (mean 0.370 ± 0.016 SE) ([Table S5](#)). Genes showed signs of saturated divergence, with estimated d_S values considerably greater than one (e.g. *atp6* $d_S = 46.8711$).

To identify potential homologues of nuclear-encoded DNA repair proteins that are targeted to the mitochondrion in higher plants and may play a role in organellar genome maintenance and size (see below), we performed blast searches of *Arabidopsis* sequences against the *O. quekettii* transcriptome and nuclear contigs. We identified a sequence in *O. quekettii* with strong similarity to the *Arabidopsis* DNA mismatch repair protein MSH1 and putative targeting to the mitochondrion (DeepLoc-1.0L: Mitochondrion 0.7151, Soluble 0.6145). Pfam and InterPro searches identified a putative specific DNA-binding GIY-YIG domain in this MSH1 homologue, as well as in potential homologues in other green algae. However, this domain lacks most of the key residues conserved among GIY-YIG family members ([Garrison & Arrizabalaga, 2009](#)). We also identified putative mitochondrion-targeted (DeepLoc-1.0L: Mitochondrion 0.9268 and 0.6334, Soluble 0.5699 and 0.6798) homologues of RECA proteins, which are also predicted to play a role in controlling mitochondrial genome maintenance in plants. Searches of the *O. quekettii* transcriptome revealed potential OSB1 homologues, predicted to be involved in homologous recombination-dependent repair, containing a central OB-fold domain but lacking a targeting signal to the mitochondrion or chloroplast (DeepLoc-1.0L: Nucleus 0.6485, Soluble 0.7227).

DISCUSSION

The inflated mitochondrial genome of *O. quekettii*, the largest Chlorophyta mitochondrial genome sequenced thus far, even larger than that of *C. lentillifera*, lends support to large mitochondrial genomes being a characteristic of the order Bryopsidales. The *O. quekettii* mitochondrial genome encodes all genes commonly found in Chlorophyta mitochondrial

genomes, including most ribosomal protein genes, which have been unevenly retained in plant mitochondria ([Palmer et al., 2000](#); [Mower, Sloan & Alverson, 2012](#)), and are missing in some Chlorophyta lineages such as the Chlorophyceae and *Pedinomonas* (e.g. [Fig. S3](#)). Surprisingly, although ribosomal proteins have been identified in other Ulvophyceae mitochondrial genomes thus far, none were identified in the mitochondrial genome of *C. lentillifera* ([Zheng et al., 2018](#)). The large size of both Bryopsidales mitochondrial genomes relative to their plastids is not typical of the Chlorophyta, where chloroplast genomes tend to be either similar size or larger and contain more noncoding DNA than their mostly compact intron-poor mitochondrial counterparts ([Leliaert et al., 2012](#)). Instead, these genomes are more typical of land plants: bloated with introns and intergenic DNA ([Leliaert et al., 2012](#); [Mower, Sloan & Alverson, 2012](#)). In fact, the Bryopsidales mitochondrial genomes are larger than those of many streptophytes, above the average range cited for non-vascular land plants but not quite as extreme as has been described in the Trachaeophytes ([Mower, Sloan & Alverson, 2012](#)). The GC content of the Bryopsidales mitochondrial genomes is also high, markedly higher than their chloroplasts. Differing nucleotide content between organelle genomes, however, is not unheard of in green algae, nor is drastic variation in organelle nucleotide content between even closely related lineages ([Smith et al., 2011](#); [Del Vasto et al., 2015](#)).

There is evidence for at least proliferation of group II introns within the *Ostreobium* lineage, something that is not uncommon in green algal chloroplast genomes ([Cremen et al., 2018](#)). However, a lack of sequence similarity for ORFs identified between their genomes, alongside the fact that their introns are not readily alignable, suggests that either the introns and associated proteins arose independently between *C. lentillifera* and *O. quekettii* before proliferating within their respective lineages, or that sequences have diverged so much that similarities are no longer recognisable. This is not unlikely given the estimated 479 million years divergence time between these lineages ([Verbruggen et al., 2009](#)). [Cremen et al. \(2018\)](#) did find some homologous ORFs, with conserved protein domains, between Bryopsidales chloroplast genomes. The lack of similarity between *Ostreobium* mitochondrial ORFs and its single plastid ORF with a putative intron splicing function ([Cremen et al., 2018](#)), as well as between introns from the organellar genomes, indicates there has been no transfer of introns between organellar genomes.

The considerable difference in their sizes and gene density as well as a lack of evidence for ORF or intron transfer between organellar compartments indicates that evolutionary forces are acting upon plastid and mitochondrial genomes differently in the Bryopsidales. The MHH ([Lynch, Koskella & Schaack, 2006](#)) and drift-barrier ([Lynch et al., 2016](#)) hypotheses emphasise the importance of mutation rate in determining genome sizes. Molecular evolution rates of the *Ostreobium* chloroplast are slow compared with other Bryopsidales; [Marcelino et al. \(2016\)](#) propose that this is due to the low light habitat of *Ostreobium*, which might reduce cases of direct sunlight-induced mutation as well as drive the evolution of slow metabolic rates and generation times with fewer mutations accumulated per unit of time. Estimates in this study indicate a slightly higher substitution rate in Bryopsidales mitochondrial genomes compared with the chloroplasts, which appears to contradict the MHH that states genomes with a higher mutation rate should

show signs of compaction, as they are under greater selective pressure to remove excess DNA that carries a mutational burden (Lynch, Koskella & Schaack, 2006). d_S values estimated from gene alignments show signs of saturated divergence, making accurate calculation of d_N/d_S unrealistic (Yang & Bielawski, 2000; Hurst, 2002). This is unsurprising due to the considerable time since the divergence of these lineages. It is not uncommon for mutation rates to vary between lineages and even between genome compartments of the same lineages by a considerably larger margin than was estimated in this study, and it can be difficult to draw direct connexions between mutation rates and genome architecture (Smith, 2016). An example is the green alga *Dunaliella salina*, which contains inflated organelle genomes, both chloroplast and mitochondrial, but has order-of-magnitude differences in mutation rates between the two compartments, with substitution rates between two strains of *D. salina* 2–13 times greater in mtDNA than ptDNA (Del Vasto et al., 2015). Differing mutation rates alone appear insufficient to explain the genome dynamics of the two Bryopsidales organellar compartments.

The higher mutation rate in the Bryopsidales mitochondrial genomes may have driven selection for more efficient DNA repair proteins, as would be predicted by the drift-barrier hypothesis (Lynch et al., 2016). It has been proposed that two distinct strategies have evolved to protect organelle genomes from the negative effects of non-homologous recombination. In animal mtDNA, elevated rates of evolution cause noncoding elements to pose a high mutational burden, leading to selection against noncoding elements, avoiding the build-up of repeats and introns (Lynch, Koskella & Schaack, 2006; Galtier, 2011). Plants appear to have efficient recombination-mediated DNA repair of coding DNA, explaining their low measured mutation rate (Odahara et al., 2009; Davila et al., 2011; Christensen, 2014). The nuclear-encoded RECA3 and MSH1 genes in plants are hypothesised to control mitochondrial genome maintenance, by preventing replication of short repeats while allowing recombination-dependent replication of longer repeats (Shedge et al., 2007). We identified a putative mitochondrion-targeted MSH1 homologue in *O. quekettii*. MSH1 encodes a protein with six conserved domains (Kowalski et al., 1999) including domain VI, a GIY-YIG homing endonuclease, which is predicted to be responsible for specific DNA-binding and suppression of homologous recombination (Fukui et al., 2018) and is specific to only the plant form of the protein (Abdelnoor et al., 2006; Shedge et al., 2007). Domain VI is absent from nuclear localised homologues in plants (MSH2–MSH6) and from the yeast MSH1 protein (Abdelnoor et al., 2006). Although InterPro and Pfam predicted a GIY-YIG domain in the *O. quekettii* MSH1 homologue, the fact that it lacks most of the residues that are typically conserved in this domain leaves its function unresolved.

Ostreobium quekettii also appears to encode mitochondrially-targeted RecA proteins. RecA recombinases in *Arabidopsis* are involved in strand exchange and the joining of paired DNA ends during homologous recombination (Kühn & Gualberto, 2012; Gualberto et al., 2014). RECA1 is chloroplast targeted, RECA2 is dual targeted to the mitochondrion and chloroplast, and RECA3 is targeted to the mitochondrion (Shedge et al., 2007). RECA3 mutations in *Arabidopsis* result in mitochondrial rearrangements similar but not identical to MSH1 mutants (Shedge et al., 2007). Odahara et al. (2009) propose that these

RecA proteins mediate homologous recombination, which is significant for suppressing short repeat-mediated genome rearrangements in plant mitochondria. They suggest that this genome stabilisation provided by RecA could allow the number of group II introns, the dominant form in the *O. quekettii* mitochondrion, to increase (Odahara et al., 2009). OSB1 is another putative component of homologous recombination-dependent repair in plant mitochondria, which is also likely involved in restricting mtDNA recombination (Kühn & Gualberto, 2012; Gualberto et al., 2014). Searches of the *O. quekettii* transcriptome revealed potential OSB1 homologues. However, these lack targeting signals to organelles that would provide evidence supporting their predicted function.

In contrast to the mitochondrial and chloroplast genomes of *V. carteri* (Smith & Lee, 2009), and those recently published for *Haematococcus lacustris* (Zhang et al., 2019), as well as the mitochondrial genomes of many land plants (Palmer et al., 2000; Mower, Sloan & Alverson, 2012), little of the expanded content of Bryopsidales mtDNA is repetitive DNA. However, most of the repeats in the *O. quekettii* mitochondrial genome are so-called ‘intermediate’ repeats (50–600 bp) (Kühn & Gualberto, 2012). Repeats of this length are associated with MSH1-induced recombination in *Arabidopsis* mitochondria, that can lead to accumulation of DNA as well as complex rearrangements (Gualberto et al., 2014). Along with error prone repair these processes might result in low numbers of alternative genome configurations, ‘mitotypes’ that could eventually increase to become the dominant form of mtDNA (Gualberto et al., 2014). This working hypothesis, based on molecular data, proposes that a recombination-associated repair process in the Bryopsidales mitochondrion has resulted in its inflated size, and encourages further study to resolve the role played by these putative mitochondrion-targeted sequences as well as to determine whether recombination is in fact occurring in Bryopsidales mitochondrial genomes at all.

It is interesting to consider whether the build-up of introns in Bryopsidales mitochondrial genomes plays a part in modifying recombination if it is in fact occurring. Introns may play a role in reducing intragenic Hill–Roberston effects, that tend to result in locally reduced N_e , by increasing recombination rate between regions of the genome (Comeron & Kreitman, 2000). Bottlenecking of genomes, such as during the production of gametes for sexual reproduction, can also result in smaller genome N_e (Neiman & Taylor, 2009), which would reduce selection and might contribute to the higher mutational load observed in Bryopsidales mitochondrial genomes. Sexual reproduction has not been observed in *O. quekettii*, however it has been described in other Bryopsidales (Morabito, Gargiulo & Genovese, 2010) and, to our knowledge, no strictly asexual lineages have been observed. Comparing relative copy numbers of organelle genomes between adult plants and gametes could provide insight into the extent of bottlenecking that occurs during reproduction in the Bryopsidales.

The mutation and recombination rates in organelles are at least in part under the control of nuclear-encoded maintenance pathways (Smith & Keeling, 2015), with organelle genomes usually lacking the genes necessary for their own DNA replication and repair (Sloan & Taylor, 2012). It has been proposed that the evolution of similar features in both plastid and mitochondrial genomes within a species, as seen in some of the Chlorophyta, is due to ‘leakage’ of nuclear-encoded proteins controlling these processes between

organellar compartments, with proteins normally targeted to one organelle also becoming targeted to the other (Smith & Keeling, 2015). The differing configurations of the two organellar genomes in *C. lentillifera* and *O. quekettii* suggest that this has not occurred in the Bryopsidales. As more nuclear genomes for Bryopsidales become available (Arimoto *et al.*, 2019), there will soon be the opportunity for a more thorough study of all three genomic compartments in the Bryopsidales, including an examination of organelle-targeted DNA maintenance machinery to further uncover the forces underpinning their divergent organelle genome sizes.

CONCLUSIONS

Marcelino *et al.* (2016) suggested that selection due to a low light environment resulted in the reduced chloroplast genome of *O. quekettii*, but it is unclear why this selection would not act against the expanded genome size in the mitochondrion. It would appear that effective population size, recombination and/or mutation rate, influenced by nuclear-encoded proteins, are different between the two genomes, leading to a reduction in the strength of selection to influence evolution of the mitochondrial genome. Ultimately, it is likely overly simplistic to attempt to find a single explanation to cover all mitochondrial genome expansion (Smith & Keeling, 2015), with the evolution of organellar genomes in the Bryopsidales and other lineages a combination of many forces and factors.

ACKNOWLEDGEMENTS

This research was supported by use of the computational facilities of the University of Melbourne High Performance Computing platform and the Nectar Research Cloud. This paper benefitted from useful discussions with Joseph Bielawski about evolution, as well as helpful comments from Geoffrey McFadden and Patrick Buerger on an earlier draft. We appreciate the thoughtful comments on the manuscript made by Jiri Neustupa and an anonymous reviewer.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Funding was provided by the Australian Research Council (DP150100705) and the University of Melbourne (Early Career Researcher grant to Heroen Verbruggen and Computational Biology Research Initiative seed funding to Heroen Verbruggen, Kathryn Holt and Linda Blackall). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Australian Research Council: DP150100705.
University of Melbourne.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Sonja I. Repetti conceived and designed the experiments, performed the experiments, analysed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Christopher J. Jackson conceived and designed the experiments, performed the experiments, analysed the data, authored or reviewed drafts of the paper, approved the final draft.
- Louise M. Judd performed the experiments, authored or reviewed drafts of the paper, approved the final draft.
- Ryan R. Wick performed the experiments, authored or reviewed drafts of the paper, approved the final draft.
- Kathryn E. Holt conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Heroen Verbruggen conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The *O. quekettii* mitochondrial genome sequence is available at GenBank: [MN514984](https://www.ncbi.nlm.nih.gov/nuclseq/MN514984).

Data Availability

The following information was supplied regarding data availability:

MAFFT alignments of ORF-lacking intron clusters identified in [Fig. S2](#) are available in a [Supplemental File](#).

Raw RNA sequencing data is available at ENA: [PRJEB35267](https://ena.ebi.ac.uk/ena/browser/view/PRJEB35267).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.8273#supplemental-information>.

REFERENCES

- Abdelnoor RV, Christensen AC, Mohammed S, Munoz-Castillo B, Moriyama H, Mackenzie SA. 2006. Mitochondrial genome dynamics in plants and animals: convergent gene fusions of a MutS homologue. *Journal of Molecular Evolution* **63**(2):165–173
DOI [10.1007/s00239-005-0226-9](https://doi.org/10.1007/s00239-005-0226-9).
- Abe T, Ikemura T, Sugahara J, Kanai A, Ohara Y, Uehara H, Kinouchi M, Kanaya S, Yamada Y, Muto A, Inokuchi H. 2010. tRNADB-CE 2011: tRNA gene database curated manually by experts. *Nucleic Acids Research* **39**(Database):D210–D213
DOI [10.1093/nar/gkq1007](https://doi.org/10.1093/nar/gkq1007).
- Alkatib S, Scharff LB, Rogalski M, Fleischmann TT, Matthes A, Seeger S, Schöttler MA, Ruf S, Bock R. 2012. The contributions of wobbling and superwobbling to the reading of the genetic code. *PLOS Genetics* **8**(11):e1003076 DOI [10.1371/journal.pgen.1003076](https://doi.org/10.1371/journal.pgen.1003076).

- Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. 2017.** DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33**(21):3387–3395 DOI [10.1093/bioinformatics/btx431](https://doi.org/10.1093/bioinformatics/btx431).
- Arimoto A, Nishitsuji K, Higa Y, Arakaki N, Hisata K, Shinzato C, Satoh N, Shoguchi E. 2019.** A siphonous macroalgal genome suggests convergent functions of homeobox genes in algae and land plants. *DNA Research* **26**(2):183–192 DOI [10.1093/dnares/dsz002](https://doi.org/10.1093/dnares/dsz002).
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012.** SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**(5):455–477 DOI [10.1089/cmb.2012.0021](https://doi.org/10.1089/cmb.2012.0021).
- Beck N, Lang B. 2010.** *MFannot, organelle genome annotation webserver*. Canada: Université de Montréal QC.
- Bolger AM, Lohse M, Usadel B. 2014.** Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15):2114–2120 DOI [10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170).
- Burger G, Gray MW, Lang BF. 2003.** Mitochondrial genomes: anything goes. *Trends in Genetics* **19**(12):709–716 DOI [10.1016/j.tig.2003.10.012](https://doi.org/10.1016/j.tig.2003.10.012).
- Carrie C, Weißenberger S, Soll J. 2016.** Plant mitochondria contain the protein translocase subunits TatB and TatC. *Journal of Cell Science* **129**(20):3935–3947 DOI [10.1242/jcs.190975](https://doi.org/10.1242/jcs.190975).
- Charlesworth B. 1994.** The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genetics Research* **63**(3):213–227 DOI [10.1017/S0016672300032365](https://doi.org/10.1017/S0016672300032365).
- Christensen AC. 2014.** Genes and junk in plant mitochondria—repair mechanisms and selection. *Genome Biology and Evolution* **6**(6):1448–1453 DOI [10.1093/gbe/evu115](https://doi.org/10.1093/gbe/evu115).
- Cocquyt E, Verbruggen H, Leliaert F, De Clerck O. 2010.** Evolution and cytological diversification of the green seaweeds (Ulvophyceae). *Molecular Biology and Evolution* **27**(9):2052–2061 DOI [10.1093/molbev/msq091](https://doi.org/10.1093/molbev/msq091).
- Comeron JM, Kreitman M. 2000.** The correlation between intron length and recombination in *Drosophila*: dynamic equilibrium between mutational and selective forces. *Genetics* **156**:1175–1190.
- Comeron JM, Williford A, Kliman RM. 2008.** The Hill–Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. *Heredity* **100**(1):19–31 DOI [10.1038/sj.hdy.6801059](https://doi.org/10.1038/sj.hdy.6801059).
- Cremen MCM, Huisman JM, Marcelino VR, Verbruggen H. 2016.** Taxonomic revision of *Halimeda* (Bryopsidales, Chlorophyta) in south-western Australia. *Australian Systematic Botany* **29**(1):41–54 DOI [10.1071/SB15043](https://doi.org/10.1071/SB15043).
- Cremen MCM, Leliaert F, Marcelino VR, Verbruggen H. 2018.** Large diversity of nonstandard genes and dynamic evolution of chloroplast genomes in siphonous green algae (Bryopsidales, Chlorophyta). *Genome Biology and Evolution* **10**(4):1048–1061 DOI [10.1093/gbe/evy063](https://doi.org/10.1093/gbe/evy063).
- Cutter AD, Payseur BA. 2013.** Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics* **14**(4):262–274 DOI [10.1038/nrg3425](https://doi.org/10.1038/nrg3425).
- Davila JI, Arrieta-Montiel MP, Wamboldt Y, Cao J, Hagemann J, Shedge V, Xu Y-Z, Weigel D, Mackenzie SA. 2011.** Double-strand break repair processes drive evolution of the mitochondrial genome in *Arabidopsis*. *BMC Biology* **9**(1):64 DOI [10.1186/1741-7007-9-64](https://doi.org/10.1186/1741-7007-9-64).
- De Vries J, Habicht J, Woehle C, Huang C, Christa G, Wägele H, Nickelsen J, Martin WF, Gould SB. 2013.** Is ftsH the key to plastid longevity in sacoglossan slugs? *Genome Biology and Evolution* **5**(12):2540–2548 DOI [10.1093/gbe/evt205](https://doi.org/10.1093/gbe/evt205).

- Del Cortona A, Leliaert F, Bogaert KA, Turmel M, Boedeker C, Janoušková J, Lopez-Bautista JM, Verbruggen H, Vandepoele K, De Clerck O. 2017. The plastid genome in cladophorales green algae is encoded by hairpin chromosomes. *Current Biology* 27(24):3771–3782.e6 DOI 10.1016/j.cub.2017.11.004.
- Del Vasto M, Figueroa-Martinez F, Featherston J, Gonzalez MA, Reyes-Prieto A, Durand PM, Smith DR. 2015. Massive and widespread organelle genomic expansion in the green algal genus *Dunaliella*. *Genome Biology and Evolution* 7(3):656–663 DOI 10.1093/gbe/evv027.
- Fang L, Leliaert F, Zhang ZH, Penny D, Zhong BJ. 2017. Evolution of the chlorophyta: insights from chloroplast phylogenomic analyses. *Journal of Systematics and Evolution* 55(4):322–332 DOI 10.1111/jse.12248.
- Felsenstein J. 2005. *PHYLIP (Phylogeny inference package)*. Version 3.6. Seattle: Department of Genome Sciences, University of Washington.
- Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y, Dosztányi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto S C E, Wu CH, Xenarios I, Yeh L-S, Young S-Y, Mitchell AL. 2017. InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Research* 45(D1):D190–D199 DOI 10.1093/nar/gkw1107.
- Frickey T, Lupas A. 2004. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20(18):3702–3704 DOI 10.1093/bioinformatics/bth444.
- Fučíková K, Leliaert F, Cooper ED, Škaloud P, D'hondt S, De Clerck O, Gurgel CF, Lewis LA, Lewis PO, Lopez-Bautista JM. 2014. New phylogenetic hypotheses for the core chlorophyta based on chloroplast sequence data. *Frontiers in Ecology and Evolution* 2:63.
- Fukui K, Harada A, Wakamatsu T, Minobe A, Ohshita K, Ashiuchi M, Yano T. 2018. The GIY-YIG endonuclease domain of *Arabidopsis* MutS homolog 1 specifically binds to branched DNA structures. *FEBS Letters* 592(24):4066–4077 DOI 10.1002/1873-3468.13279.
- Galtier N. 2011. The intriguing evolutionary dynamics of plant mitochondrial DNA. *BMC Biology* 9(1):61 DOI 10.1186/1741-7007-9-61.
- Garrison EM, Arrizabalaga G. 2009. Disruption of a mitochondrial MutS DNA repair enzyme homologue confers drug resistance in the parasite *Toxoplasma gondii*. *Molecular Microbiology* 72(2):425–441 DOI 10.1111/j.1365-2958.2009.06655.x.
- Gelfand Y, Rodriguez A, Benson G. 2007. TRDB—the tandem repeats database. *Nucleic Acids Research* 35(Database):D80–D87 DOI 10.1093/nar/gkl1013.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29(7):644–652 DOI 10.1038/nbt.1883.
- Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Research* 31(1):439–441 DOI 10.1093/nar/gkg006.
- Gualberto JM, Milesina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A. 2014. The plant mitochondrial genome: dynamics and maintenance. *Biochimie* 100:107–120 DOI 10.1016/j.biochi.2013.09.016.
- Hausner G. 2012. Introns, mobile elements, and plasmids. In: Bullerwell CE, ed. *Organelle Genetics*. Berlin: Springer, 329–357.

- Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genetics Research* 8(3):269–294 DOI 10.1017/S0016672300010156.
- Hurst LD. 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends in Genetics* 18(9):486–487 DOI 10.1016/S0168-9525(02)02722-1.
- Jukes TH, Cantor CR. 1969. Evolution of protein molecules. *Mammalian Protein Metabolism* 3:132.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649 DOI 10.1093/bioinformatics/bts199.
- Kinouchi M, Kurokawa K. 2006. [Special issue: fact databases and freewares] tRNAfinder: a software system to find all tRNA genes in the DNA sequence based on the cloverleaf secondary structure. *Journal of Computer Aided Chemistry* 7:116–124 DOI 10.2751/jcac.7.116.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Research* 27(5):722–736 DOI 10.1101/gr.215087.116.
- Kowalski JC, Belfort M, Stapleton MA, Holpert M, Dansereau JT, Pietrokovski S, Baxter SM, Derbyshire V. 1999. Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Research* 27(10):2115–2125 DOI 10.1093/nar/27.10.2115.
- Krasovec M, Eyre-Walker A, Sanchez-Ferandin S, Piganeau G. 2017. Spontaneous mutation rate in the smallest photosynthetic eukaryotes. *Molecular Biology and Evolution* 34(7):1770–1779 DOI 10.1093/molbev/msx119.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Research* 19(9):1639–1645 DOI 10.1101/gr.092759.109.
- Kühn K, Gualberto JM. 2012. Recombination in the stability, repair and evolution of the mitochondrial genome. In: Maréchal-Drouard L, ed. *Advances in Botanical Research* 63. Oxford: Academic Press, 215–252.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35(6):1547–1549 DOI 10.1093/molbev/msy096.
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* 29(22):4633–4642 DOI 10.1093/nar/29.22.4633.
- Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research* 35(9):3100–3108 DOI 10.1093/nar/gkm160.
- Lambowitz AM, Belfort M. 1993. Introns as mobile genetic elements. *Annual Review of Biochemistry* 62(1):587–622 DOI 10.1146/annurev.bi.62.070193.003103.
- Lambowitz AM, Zimmerly S. 2011. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harbor Perspectives in Biology* 3(8):a003616 DOI 10.1101/cshperspect.a003616.
- Lang BF, Laforest M-J, Burger G. 2007. Mitochondrial introns: a critical view. *Trends in Genetics* 23(3):119–125 DOI 10.1016/j.tig.2007.01.006.
- Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research* 32(1):11–16 DOI 10.1093/nar/gkh152.

- Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, De Clerck O. 2012. Phylogeny and molecular evolution of the green algae. *Critical Reviews in Plant Sciences* 31(1):1–46 DOI 10.1080/07352689.2011.615705.
- Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research* 44(W1):W54–W57 DOI 10.1093/nar/gkw413.
- Lynch M. 2006. Streamlining and simplification of microbial genome architecture. *Annual Review of Microbiology* 60(1):327–349 DOI 10.1146/annurev.micro.60.080805.142300.
- Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics* 17(11):704–714 DOI 10.1038/nrg.2016.104.
- Lynch M, Koskella B, Schaack S. 2006. Mutation pressure and the evolution of organelle genomic architecture. *Science* 311(5768):1727–1730 DOI 10.1126/science.1118884.
- Lynch M, Walsh B. 2007. *The origins of genome architecture* 98. Sunderland: Sinauer Associates.
- Magnusson SH, Fine M, Kühl M. 2007. Light microclimate of endolithic phototrophs in the scleractinian corals *Montipora monasteriata* and *Porites cylindrica*. *Marine Ecology Progress Series* 332:119–128 DOI 10.3354/meps332119.
- Marcelino VR, Cremen MCM, Jackson CJ, Larkum AAW, Verbruggen H. 2016. Evolutionary dynamics of chloroplast genomes in low light: a case study of the endolithic green alga *Ostreobium quekettii*. *Genome Biology and Evolution* 8(9):2939–2951 DOI 10.1093/gbe/evw206.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17(1):10–12 DOI 10.14806/ej.17.1.200.
- Mine I, Sekida S, Okuda K. 2015. Cell wall and cell growth characteristics of giant-celled algae. *Phycological Research* 63(2):77–84 DOI 10.1111/pre.12088.
- Morabito M, Gargiulo G, Genovese G. 2010. A review of life history pathways in *Bryopsis* species. *Atti della Accademia Peloritana dei Pericolanti-Classe di Scienze Fisiche, Matematiche e Naturali LXXXVIII(2):C1A1002005* DOI 10.1478/C1A1002005.
- Mower JP, Sloan DB, Alverson AJ. 2012. Plant mitochondrial genome diversity: the genomics revolution. In: Wendel J, Greilhuber J, Dolezel J, Leitch IJ, eds. *Plant Genome Diversity*. Vol. 1. Wien: Springer, 123–144.
- Neiman M, Taylor DR. 2009. The causes of mutation accumulation in mitochondrial genomes. *Proceedings of the Royal Society B: Biological Sciences* 276(1660):1201–1209 DOI 10.1098/rspb.2008.1758.
- Ness RW, Kraemer SA, Colegrave N, Keightley PD. 2016. Direct estimate of the spontaneous mutation rate uncovers the effects of drift and recombination in the *Chlamydomonas reinhardtii* plastid genome. *Molecular Biology and Evolution* 33(3):800–808 DOI 10.1093/molbev/msv272.
- Odahara M, Kuroiwa H, Kuroiwa T, Sekine Y. 2009. Suppression of repeat-mediated gross mitochondrial genome rearrangements by RecA in the moss *Physcomitrella patens*. *Plant Cell* 21(4):1182–1194 DOI 10.1105/tpc.108.064709.
- Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu Y-L, Song K. 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proceedings of the National Academy of Sciences of the United States of America* 97(13):6960–6966 DOI 10.1073/pnas.97.13.6960.
- Petrů M, Wideman J, Moore K, Alcock F, Palmer T, Doležal P. 2018. Evolution of mitochondrial TAT translocases illustrates the loss of bacterial protein transport machines in mitochondria. *BMC Biology* 16(1):141 DOI 10.1186/s12915-018-0607-3.

- Platt A, Weber CC, Liberles DA. 2018.** Protein evolution depends on multiple distinct population size parameters. *BMC Evolutionary Biology* **18**(1):17 DOI [10.1186/s12862-017-1085-x](https://doi.org/10.1186/s12862-017-1085-x).
- Pombert J-F, Otis C, Lemieux C, Turmel M. 2004.** The complete mitochondrial DNA sequence of the green alga *Pseudendoclonium akinetum* (Ulvophyceae) highlights distinctive evolutionary trends in the chlorophyta and suggests a sister-group relationship between the ulvophyceae and chlorophyceae. *Molecular Biology and Evolution* **21**(5):922–935 DOI [10.1093/molbev/msh099](https://doi.org/10.1093/molbev/msh099).
- Ricci F, Marcelino VR, Blackall L, Kühl M, Medina M, Verbruggen H. 2019.** Beneath the surface: community assembly and functions of the coral skeleton microbiome. *Microbiome* **7**:159 DOI [10.1186/s40168-019-0762-y](https://doi.org/10.1186/s40168-019-0762-y).
- Schlichter D, Kampmann H, Conrady S. 1997.** Trophic potential and photoecology of endolithic algae living within coral skeletons. *Marine Ecology* **18**(4):299–317 DOI [10.1111/j.1439-0485.1997.tb00444.x](https://doi.org/10.1111/j.1439-0485.1997.tb00444.x).
- Shedge V, Arrieta-Montiel M, Christensen AC, Mackenzie SA. 2007.** Plant mitochondrial recombination surveillance requires unusual *RecA* and *MutS* homologs. *Plant Cell* **19**(4):1251–1264 DOI [10.1105/tpc.106.048355](https://doi.org/10.1105/tpc.106.048355).
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011.** Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular Systems Biology* **7**(1):539 DOI [10.1038/msb.2011.75](https://doi.org/10.1038/msb.2011.75).
- Sloan DB, Taylor DR. 2012.** Evolutionary rate variation in organelle genomes: the role of mutational processes. In: Bullerwell CE, ed. *Organelle Genetics*. Berlin: Springer, 123–146.
- Smith DR. 2016.** The mutational hazard hypothesis of organelle genome evolution: 10 years on. *Molecular Ecology* **25**(16):3769–3775 DOI [10.1111/mec.13742](https://doi.org/10.1111/mec.13742).
- Smith DR, Burki F, Yamada T, Grimwood J, Grigoriev IV, Van Etten JL, Keeling PJ. 2011.** The GC-rich mitochondrial and plastid genomes of the green alga *Coccomyxa* give insight into the evolution of organelle DNA nucleotide landscape. *PLOS ONE* **6**(8):e23624 DOI [10.1371/journal.pone.0023624](https://doi.org/10.1371/journal.pone.0023624).
- Smith JM, Haigh J. 1974.** The hitch-hiking effect of a favourable gene. *Genetics Research* **23**(1):23–35 DOI [10.1017/S0016672300014634](https://doi.org/10.1017/S0016672300014634).
- Smith DR, Hamaji T, Olson BJSC, Durand PM, Ferris P, Michod RE, Featherston J, Nozaki H, Keeling PJ. 2013.** Organelle genome complexity scales positively with organism size in volvocine green algae. *Molecular Biology and Evolution* **30**(4):793–797 DOI [10.1093/molbev/mst002](https://doi.org/10.1093/molbev/mst002).
- Smith DR, Keeling PJ. 2015.** Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proceedings of the National Academy of Sciences of the United States of America* **112**(33):10177–10184 DOI [10.1073/pnas.1422049112](https://doi.org/10.1073/pnas.1422049112).
- Smith DR, Lee RW. 2008.** Mitochondrial genome of the colorless green alga *Polytomella capuana*: a linear molecule with an unprecedented GC content. *Molecular Biology and Evolution* **25**(3):487–496 DOI [10.1093/molbev/msm245](https://doi.org/10.1093/molbev/msm245).
- Smith D, Lee RW. 2009.** The mitochondrial and plastid genomes of *Volvox carteri*: bloated molecules rich in repetitive DNA. *BMC Genomics* **10**:132 DOI [10.1186/1471-2164-10-132](https://doi.org/10.1186/1471-2164-10-132).
- Smith DR, Lee RW. 2010.** Low nucleotide diversity for the expanded organelle and nuclear genomes of *Volvox carteri* supports the mutational-hazard hypothesis. *Molecular Biology and Evolution* **27**(10):2244–2256 DOI [10.1093/molbev/msq110](https://doi.org/10.1093/molbev/msq110).
- Smith DR, Lee RW, Cushman JC, Magnuson JK, Tran D, Polle JEW. 2010.** The *Dunaliella salina* organelle genomes: large sequences, inflated with intronic and intergenic DNA. *BMC Plant Biology* **10**(1):83 DOI [10.1186/1471-2229-10-83](https://doi.org/10.1186/1471-2229-10-83).

- Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* 34(Suppl. 2):W609–W612 DOI 10.1093/nar/gkl315.
- Tribollet A. 2008. The boring microflora in modern coral reef ecosystems: a review of its roles. In: Wisshak M, Tapanila L, eds. *Current Developments in Bioerosion*. Berlin: Springer, 67–94.
- Verbruggen H, Ashworth M, LoDuca ST, Vlaeminck C, Cocquyt E, Sauvage T, Zechman FW, Littler DS, Littler MM, Leliaert F. 2009. A multi-locus time-calibrated phylogeny of the siphonous green algae. *Molecular Phylogenetics and Evolution* 50(3):642–653 DOI 10.1016/j.ympev.2008.12.018.
- Verbruggen H, Marcelino VR, Guiry MD, Cremen MCM, Jackson CJ. 2017. Phylogenetic position of the coral symbiont *Ostreobium* (Ulvophyceae) inferred from chloroplast genome data. *Journal of Phycology* 53(4):790–803 DOI 10.1111/jpy.12540.
- Verbruggen H, Tribollet A. 2011. Boring algae. *Current Biology* 21(21):R876–R877 DOI 10.1016/j.cub.2011.09.014.
- Vroom PS, Smith CM. 2001. The challenge of siphonous green algae. *American Scientist* 89(6):525–531 DOI 10.1511/2001.6.524.
- Vroom PS, Smith CM. 2003. Life without cells. *Biologist* 50:222–226.
- Wilhelm C, Jakob T. 2006. Uphill energy transfer from long-wavelength absorbing chlorophylls to PS II in *Ostreobium* sp. is functional in carbon assimilation. *Photosynthesis Research* 87(3):323 DOI 10.1007/s11120-005-9002-3.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20(17):3252–3255 DOI 10.1093/bioinformatics/bth352.
- Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution* 15(12):496–503 DOI 10.1016/S0169-5347(00)01994-7.
- Yurina N, Odintsova M. 2016. Mitochondrial genome structure of photosynthetic eukaryotes. *Biochemistry (Moscow)* 81(2):101–113 DOI 10.1134/S0006297916020048.
- Zhang X, Bauman N, Brown R, Richardson TH, Akella S, Hann E, Morey R, Smith DR. 2019. The mitochondrial and chloroplast genomes of the green alga *Haematococcus* are made up of nearly identical repetitive sequences. *Current Biology* 29(15):R736–R737 DOI 10.1016/j.cub.2019.06.040.
- Zheng F, Liu H, Jiang M, Xu Z, Wang Z, Wang C, Du F, Shen Z, Wang B. 2018. The complete mitochondrial genome of the *Caulerpa lentillifera* (Ulvophyceae, Chlorophyta): sequence, genome content, organization structure and phylogenetic consideration. *Gene* 673:225–238 DOI 10.1016/j.gene.2018.06.050.
- Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29(21):2669–2677 DOI 10.1093/bioinformatics/btt476.
- Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely Reimplemented MPI bioinformatics toolkit with a new HHpred server at its Core. *Journal of Molecular Biology* 430(15):2237–2243 DOI 10.1016/j.jmb.2017.12.007.