Article

Current Biology

Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes

Highlights

- Cladophorales plastomes are ancestrally intact, circular, and reduced
- Plastome gene loss and mutation rate elevation preceded fragmentation
- Fragmentation resulted in extremely heterogeneous minicircular chromosomes
- UGA codons were reassigned concurrently in fragmented and non-fragmented lineages

Authors

Saelin Bjornson, Trevor T. Bringloe, Kavitha Uthanumallian, ..., Sebastian Duchene, Jacob L. Steenwyk, Heroen Verbruggen

Correspondence

skbjornson@hotmail.ca

In brief

To investigate the origins of single-stranded hairpin plastid chromosomes of *Boodlea*, Bjornson et al. employ long-read sequencing of plastomes from three Cladophorales families. Their findings show stepwise changes in plastome characteristics, including gene loss, genome fragmentation, mutation rate elevation, sequence heterogeneity, GC content elevation, codon reassignment, and chromosome topological changes.

Bjornson et al., 2026, Current Biology *36*, 1–16
April 6, 2026
https://doi.org/10.1016/j.cub.2026.02.038

CellPress

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

Current Biology

CellPress
OPEN ACCESS

Article

Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes

Saelin Bjornson,[1,8,*] Trevor T. Bringloe,[2] Kavitha Uthanumallian,[1] Ricardo De Paoli-Iseppi,[3] Michael B. Clark,[3] Chiela M.C. Cremen,[1] Riyad Hossen,[1] Owen McGinley,[1] Tara Jalali,[1] John West,[1] Frederik Leliaert,[4] Richard Wetherbee,[1] Olivier De Clerck,[5] Sebastian Duchene,[6] Jacob L. Steenwyk,[7] and Heroen Verbruggen[1]

[1]School of Biosciences, University of Melbourne, Melbourne, VIC 3010, Australia
[2]Department of Fisheries and Oceans, Government of Canada, Ottawa, ON K1A0E6, Canada
[3]Department of Anatomy and Physiology, University of Melbourne, Melbourne, VIC 3010, Australia
[4]Meise Botanic Garden, Meise 1860, Belgium
[5]Department of Biology, Phycology Research Group, Ghent University, 9000 Ghent, Belgium
[6]Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, VIC 3010, Australia
[7]Howard Hughes Medical Institute and the Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720-1234, USA
[8]Lead contact
*Correspondence: skbjornson@hotmail.ca
https://doi.org/10.1016/j.cub.2026.02.038

SUMMARY

While organellar genomes are typically a single circular molecule, some have undergone extreme evolutionary changes, such as fragmentation. Among plastid genomes, extensive fragmentation is rare but has been observed in the green alga *Boodlea composita* (Cladophorales), wherein single-stranded chromosomes adopt an unusual hairpin structure, stabilized by inverted repeats. This plastome is also highly reduced in coding content, containing only 21 protein-coding genes that are highly divergent in sequence, yet the plastome comprises a numerous and heterogeneous chromosome population. The lack of plastomes among close relatives has stymied our understanding of the evolutionary origins of *Boodlea*'s unique plastome characteristics. Here, we report plastomes from three additional Cladophorales families. We show that elevated mutation rates and a massive reduction in coding content preceded fragmentation and that fragmentation resulted in a highly heterogeneous population with hundreds of minicircular chromosomes and a concomitant increase in GC content. Long inverted repeats are absent in minicircular chromosomes, suggesting the repetitive sequences that facilitate the formation of hairpin structures are unique to the *Boodlea* clade. Codon reassignments also occurred concurrently among fragmented and non-fragmented Cladophorales plastomes. Our findings document a clear stepwise progression of plastid topologies and sequence characteristics in Cladophorales. This assists in our understanding of the evolution of fragmented organellar genomes in general, as well as the limits to coding capacity reduction and sequence divergence able to be withstood by a functioning plastid.

INTRODUCTION

Eukaryotic organelles are derived from ancient endosymbiosis with bacteria, followed by extensive gene transfer to the host nuclear genome and gene loss from the organelle. Mitochondrial genomes have undergone more drastic evolutionary fates when compared with plastid genomes, including genome fragmentation in several lineages. The extent of mitogenome fragmentation can vary. For instance, mitogenomes can break into two or three autonomous chromosomes, as observed in various green algae,[1] plants,[2] animals,[3] and fungi.[4] Other lineages have undergone more considerable mitogenome fragmentation, such as the linear chromosomes found in metazoan lineages Cnidaria and lice,[5,6] or the minicircular chromosomes found in Stylonematophyceae of the red algae[7] and the holoparasitic plant *Rhopalocnemis*.[8] In extreme cases, many tens-to-hundreds of mitogenome fragments exist, as observed in Euglenozoa,[9] the Ichthyosporean parasite *Amoebidium*,[10] and species of the angiosperm *Silene*.[11] By contrast, minor fragmentation of plastid genomes is reported in the green alga *Koshicola*,[12] while more extensive fragmentation has only been observed in dinoflagellates, whose plastomes are composed of minicircles,[13] and in the green alga *Boodlea composita* (class: Ulvophyceae; order: Cladophorales; family: Boodleaceae).[14]

The plastome of *Boodlea* presents a case of extreme genome evolution. *Boodlea* chromosomes are approximately 1–7 kilobases (kb) in length, and only one or two genes are found on a single chromosome. The overall coding content is also highly

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

reduced, containing only 21 unique protein-coding genes, whereas Chlorophyta plastomes typically encode ~60–130.[15,16] These plastid genes are also extremely divergent in sequence and have reassigned the UGA stop codon.[14] The overall chromosome population is very heterogeneous, with many pseudogene copies and non-coding chromosomes present, and has a very elevated GC content of 57%, whereas ~25%–40% is a common range in Chlorophyta.[15,16] While most of these traits are documented in other fragmented organellar genomes, the *Boodlea* plastome is unique in that it is the only sequenced genome containing entirely single-stranded, hairpin chromosomes, a topology likely present in other families of the *Siphonocladus* clade to which Boodleaceae belongs.[17] The hairpin structure is enabled by the presence of long inverted repeats (IRs) that extend the length of most chromosomes.

Despite *Boodlea*'s novelty, no other plastomes have been sequenced from Cladophorales, which represent over 700 million years of evolution,[18] and therefore the origins of *Boodlea*'s non-canonical plastome architecture, as well as the progression of gene loss and substitution rate elevation, remain unknown. To illuminate the evolutionary history of the *Boodlea* plastid genome, we sequenced the plastid genomes of three other Cladophorales families—*Chaetomorpha tokyoensis* of the Cladophoraceae, *Aegagropila linnaei* of the Pithophoraceae, and two *Okellya* species of the Okellyaceae.[19,20] We report a stepwise order of events pertaining to many characteristics of the *Boodlea* plastome, including acceleration of mutation rate, gene loss, fragmentation, topology transitions, genetic code changes, chromosome heteroplasmy, GC content increase, and emergence of IRs. These findings expand our understanding of the evolutionary forces driving drastic plastome changes.

RESULTS

Minicircles are an evolutionary intermediate to hairpin chromosomes

Nanopore sequencing was conducted for the plastid genome of *Chaetomorpha* of the Cladophoraceae, the sister family to the *Siphonocladus* clade containing *Boodlea* (consisting of Boodleaceae and other families), while PacBio HiFi sequencing was conducted for *Aegagropila* of the Pithophoraceae, sister to both clades[19] (Figures 1A and S1). Targeted assembly of plastomes confirmed these plastomes are highly fragmented, and raw assemblies of most plastid chromosomes are over-assembled, suggesting circular molecules. Circularity was confirmed by extracting non-overlapping units from assemblies and mapping long reads to side-by-side unit concatenations, showing consistent read coverage across the entire concatenation and an absence of reads with alignments longer than a unit length (Figure S2). The presence of minicircular plastid chromosomes in both *Chaetomorpha* and *Aegagropila* reveals these to be the likely evolutionary intermediate to hairpins of *Boodlea*.

The *Boodlea* plastome is represented by 34 assembled and annotated chromosomes, encoding 21 protein-coding genes and the 16S rRNA gene. This gene repertoire was recovered in *Chaetomorpha*, while in *Aegagropila*, all but *psbJ* were found, and only pseudogenes of *petD* were detected, in addition to both 16S and 23S rRNA genes. We also identified 19 unique tRNA genes in *Aegagropila* and 5 in *Chaetomorpha*, while none were found in a search of published *Boodlea* chromosomes. Like *Boodlea*, each chromosome encodes only one or two genes or gene fragments (Figure 1B). While long IRs characterize *Boodlea* chromosomes, extensive repeats are mostly absent from minicircular chromosomes (Figure 1B). However, some chromosomes, particularly in *Aegagropila*, have small, IRs spaced apart from each other or small palindromes, while *Chaetomorpha* more often contains small direct repeats and rarely inverted ones (Figure S3).

Extensive genome reduction, gene loss, and elevated substitution rates preceded fragmentation in Cladophorales plastid genomes

To gain insight into the evolution of minicircle plastomes, we sequenced two plastid genomes of the "first-branching" family of the Cladophorales, the Okellyaceae—*Okellya marina*[20] using both Nanopore and Illumina platforms, and a novel species, *Okellya* sp. SL, with Illumina. *O. marina* assembled as a single, circular-mapping molecule, while the *Okellya* sp. SL short-read assembly indicated a likely circular configuration of the plastome is present (Figures 1B, 2A, and S4). The *O. marina* plastome contains a 4.8 kb region of complex IRs centered around a palindrome (Figure 2B). This was not found in the circular configuration of *Okellya* sp. SL, however, has a similar repetitive region present at one end of the raw assembly, and therefore a repeat region is likely present, but its sequence and placement within the genome cannot be resolved (Figure S4). Some "standard" plastomes that assemble as a single circle are now known to also consist of complex multimeric branched molecules or other recombinant structures,[21,22] and it is likely that *Okellya* plastomes also have multiple and potentially non-circular configurations, one of which has no repeat region in *Okellya* sp. SL. However, in *O. marina*, mapping long reads to the plastome did not find evidence of molecules without the repeat region, and an assembly with only short reads produced repeats at both ends, unlike *Okellya* sp. SL. Regardless, *Okellya* plastomes reveal that fragmentation is not ancestral to the Cladophorales and that these plastomes are highly reduced, at only 29.2 kb in *O. marina* and 23.1 kb in *Okellya* sp. SL, taking the circular representation. In comparison, Ulvophyceae plastomes range from 82 to 408 kb.[15,16]

Okellya plastomes encode all 21 protein-coding genes found in other Cladophorales plastomes and both 16S and 23S rRNA genes. Three putative long open reading frames (ORFs) were also found in both plastomes with no detectable homology. *O. marina* contains ORFs that are an estimated 332 amino acids (aa), 187 aa, and 114 aa in length, while those in *Okellya* sp. SL are 356 aa, 201 aa, and 124 aa in length. The longest ORFs are immediately adjacent to repeat regions, which in *Okellya* sp. SL is not included in the circular representation (Figure S4). ORF114 and ORF124 as well as ORF187 and ORF201 are homologous, as they align with 31% and 38% pairwise identities, respectively, while the two largest ORFs on each genome have no sequence similarity.

Additionally, 17 tRNA genes were found on *O. marina* and 15 on *Okellya* sp. SL. With the tRNAs identified in fragmented plastome lineages, the ancestral Cladophorales plastome would have had at least 24 tRNAs for 19 aa, and 10 tRNAs (10 aa) are

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

CellPress
OPEN ACCESS

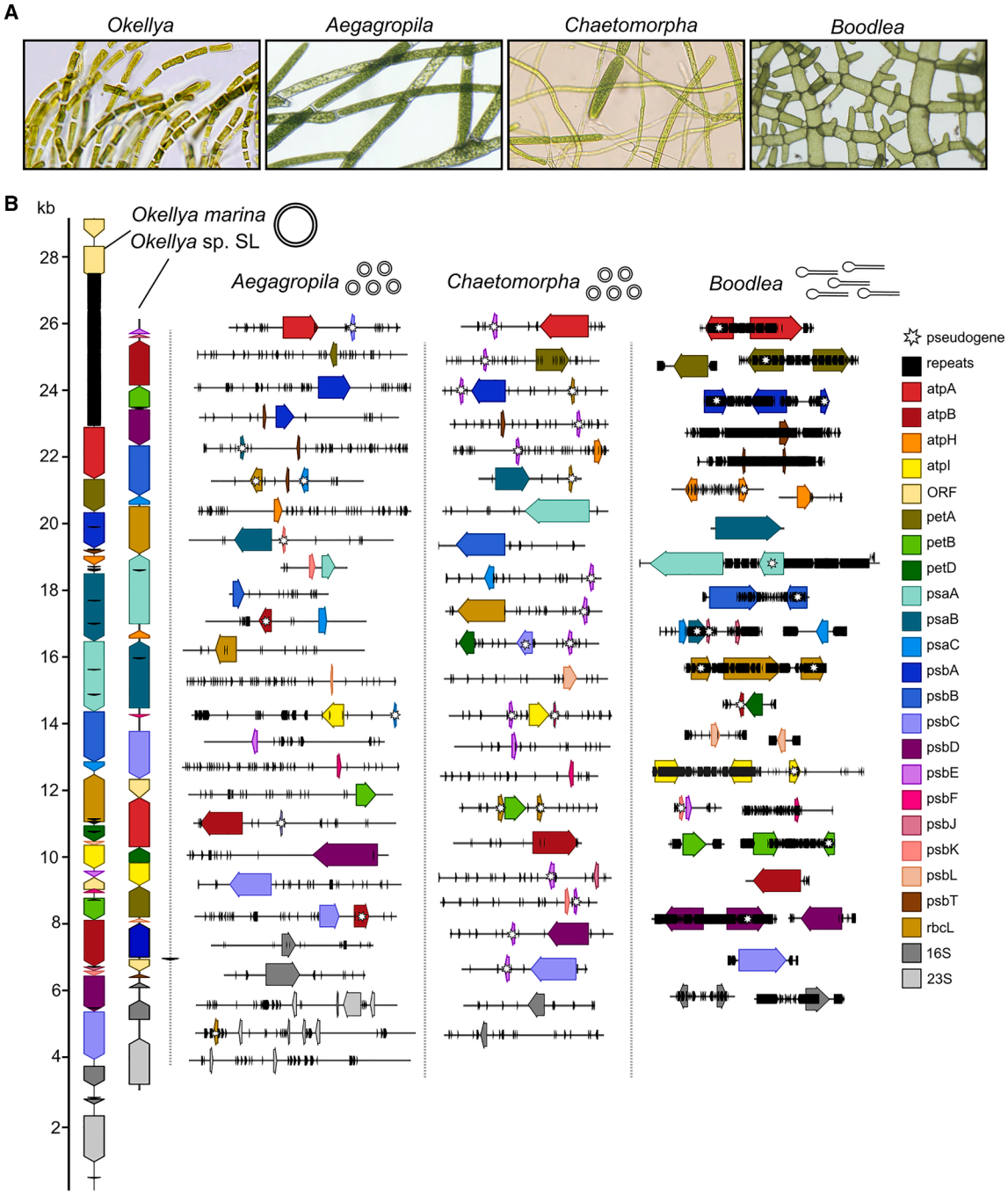

Figure 1. Evolution of Cladophorales plastid genome structure

(A) Light micrographs of our study organisms *Okellya*, *Aegagropila*, *Chaetomorpha*, and *Boodlea*. See Figure S1 for species confirmation.

(B) Schematic of Cladophorales plastid chromosomes, with chromosome topology (single circular molecule, multiple minicircles, or multiple hairpin chromosomes) indicated as symbols beside names. All chromosomes are to scale, with a scale bar on the left. The gene position and orientation of minicircular chromosomes are random and not in reference to any feature on chromosomes.

See also Figures S2 and S3, Table S1, and Data S1.

shared between the Okellyeacea and fragmented lineages. At least 17 tRNAs (16 aa) would have been present in the ancestral Okellyaceae genome, while at least 20 tRNAs (16 aa) would have been present in the ancestor of fragmented genomes. There is no correlation with tRNA anticodon GC content and retention in higher GC plastomes.

In *Okellya* plastomes, both 16S and 23S genes contain large internal deletions, and 16S genes show evidence of

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

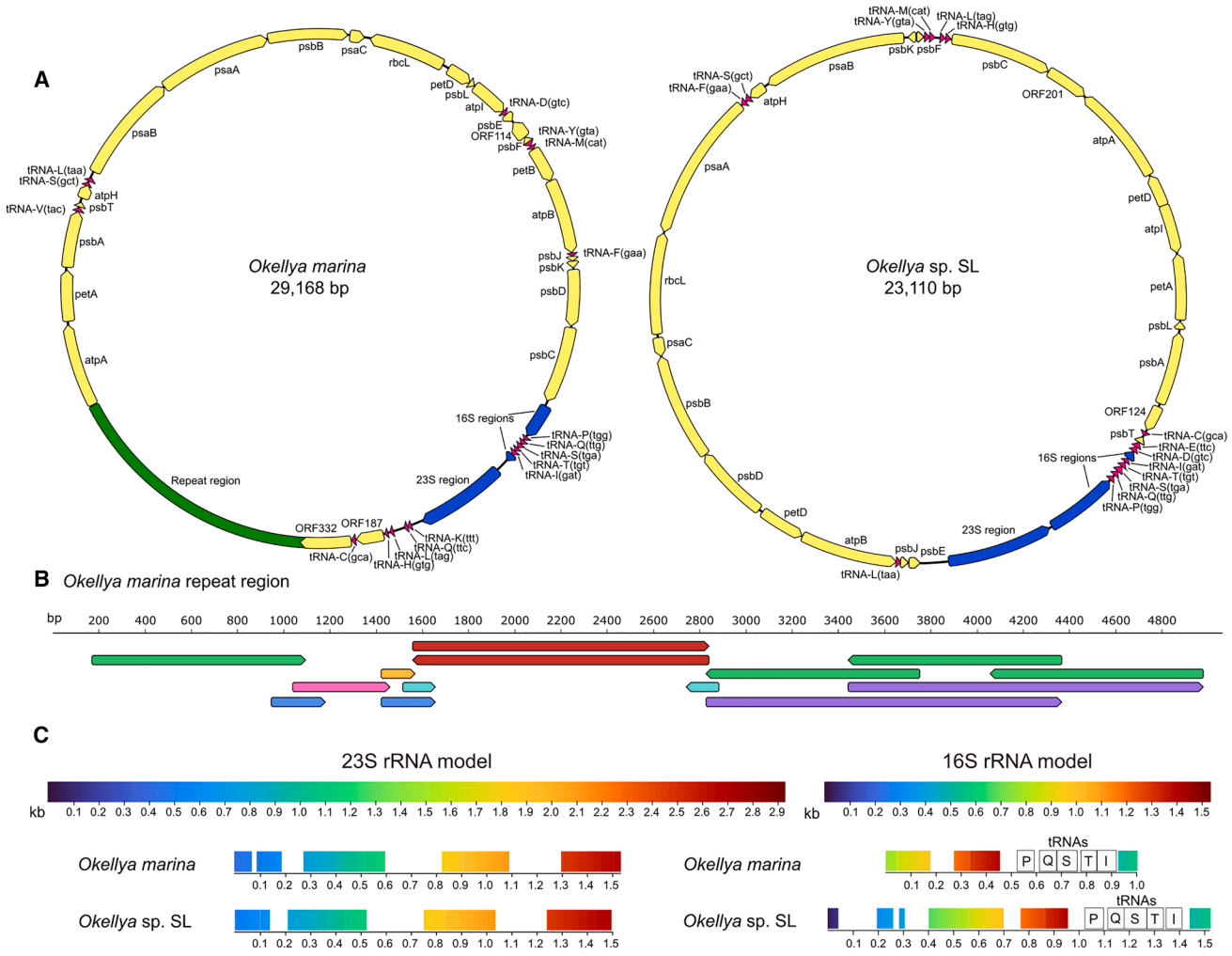

Figure 2. Details of *Okellya* plastid genomes

(A) Circular representation of *Okellya* plastid genome assemblies, indicating coding regions, tRNA genes, 16S and 23S genes, and the repeat region in *O. marina*. Genomes are not to scale.

See also Figure S4 for details of *Okellya* sp. SL plastome assembly.

(B) Details of the repeat region of the *Okellya marina* plastid. Colors indicate matching repeats. Internal, more complex repeats within larger ones are not shown.

(C) Details of *Okellya* plastid rRNA genes. Lengths of the 23S and 16S rRNA Infernal models (from Rfam RF02541 and RF00017 alignments) are shown above. Below are indicated regions of alignments on *Okellya* plastids. tRNAs are labeled by cognate aa.

rearrangements, but rRNA genes are very similar between *Okellya* species (Figure 2C). The gene order of the two *Okellya* plastids differs markedly, suggesting a high degree of recombination in this lineage (Figure 1B).

Using representative genes from minicircular chromosomes (STAR Methods), a 17-plastid gene phylogeny highlights the extremely elevated substitution rates in Cladophorales, which, by comparison of branch lengths, are particularly elevated in Okellyaceae (Figure 3A). Both dN and dS values are highly elevated in Cladophorales compared with other green algae plastids, and ratios indicate relaxation of purifying selection (Figure 3B). With this, *Okellya* plastids reveal that elevated substitution rates and extreme streamlining of coding content occurred prior to fragmentation into minicircles in other Cladophorales.

In all sequenced genomes presented here, the GC content of coding and non-coding regions was similar. We compared the GC content of plastid coding regions available from other Cladophorales, grouped by major clades in branching order, indicating that the GC content of plastid genomes shows a phylogenetic pattern across Cladophorales lineages, with later-diverging clades having higher GC content than earlier-diverging ones (Figure 3C). *Okellya* plastid coding sequences have an average GC content of 34%, whereas averages of 39%, 51%, and 55% are found in Pithophoraceae, Cladophoraceae, and the *Siphonocladus* clade (Boodleaceae and other families), respectively. GC content does not appear to be directly tied to genome architecture, as minicircular genomes are found in both Pithophoraceae and Cladophoraceae.

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

CellPress
OPEN ACCESS

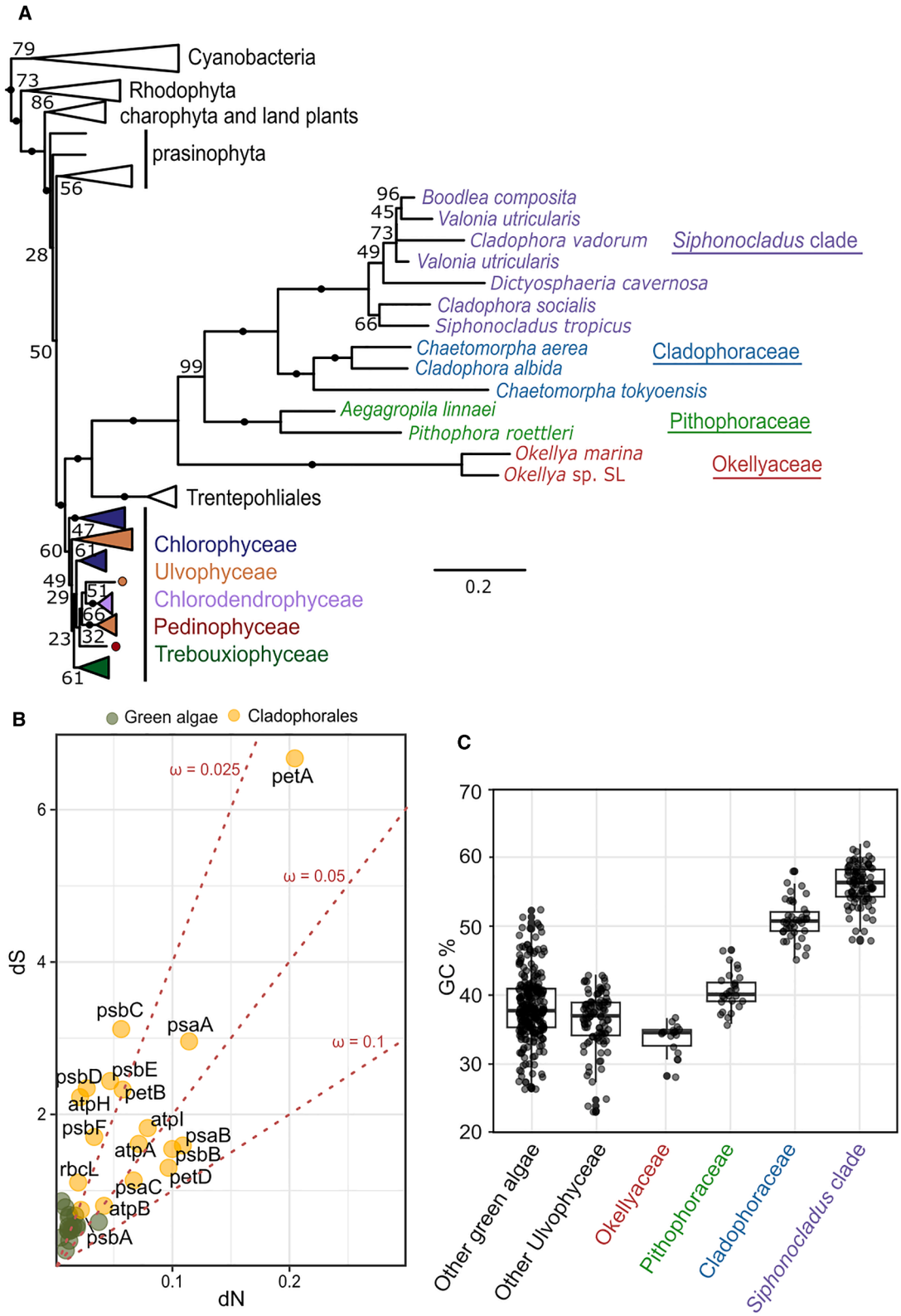

Figure 3. Divergence of coding sequences and stepwise elevation of GC content across Cladophorales plastomes

(A) Maximum-likelihood phylogeny of 17 plastid genes, showing the accelerated evolution in Cladophorales compared with other green algae. Nodes with circles represent 100% bootstrap support out of 200 replicates. The scale bar denotes aa changes per site.

(legend continued on next page)

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

Minicircular genomes are highly heterogeneous

Hundreds of minicircular chromosomes were assembled for both *Aegagropila* and *Chaetomorpha* plastomes. The diversity of aligned reads, particularly HiFi reads in *Aegagropila*, which are 99.9% accurate, suggests that assembled chromosomes should be regarded as consensus sequences of many variants. When raw assemblies are clustered at 90% nucleotide identity, 743 final chromosomes in *Aegagropila* and 243 in *Chaetomorpha* were recovered, referred to hereafter as "nr90 chromosomes" (STAR Methods; Data S1A) (Figure 4A). To estimate the proportion of total plastid-derived reads represented in *Aegagropila* HiFi assemblies, we performed sequence binning of all reads by k-mer composition and coverage. This created a bin that contained 95% of plastome-mapping reads, which comprised 73% of all reads within this bin. Likely due to limited coverage of this large sequence population, in both species a minority of nr90 chromosomes did not assemble as a closed circle, and their exact topology could not be confirmed. Of those that did assemble circularly, chromosomes are on average larger in *Aegagropila* (mean 6.3 kb) than in *Chaetomorpha* (mean 4.6 kb), which in turn are larger than the linear chromosomes of *Boodlea*, where the total of 136 plastid contigs assembled have a mean of 2 kb, while the 34 published and annotated contigs have a mean of 2.6 kb.[14]

Alignment of Illumina reads to individual minicircular chromosomes highlights that many polymorphisms (SNVs and small indels) exist in both coding and non-coding regions (Figure 4B). Among both *Aegagropila* and *Chaetomorpha* plastomes, potentially functional variants of genes are found (those with a few aa differences, small in-frame insertions/deletions, or extended ORFs), while many more non-functional copies of each gene are present, such as gene fragments and genes with internal stop codons and frame-shift mutations, many of which were verified in sequence by inspection of long-read alignments and are therefore not assembly artifacts (Figures 4C and S5). To assist in determining which gene copies are functional, we assembled RNA transcripts *de novo*. For some genes in *Aegagropila*, the representative copy (the most conserved gene copy determined by BLASTX to Chlorophyta proteins) was present full-length in transcripts and found in individual reads but was not present in assembled chromosomes and was unable to be assembled by targeted assembly, likely due to low sequencing coverage of these specific copies (noted in Table S1). Complete rRNA genes were also not found but are instead found as diverse fragments on separate chromosomes (Figure 4D). We searched for evidence of trans-splicing in transcriptomes but did not find more complete rRNA sequences. In total, genes or pseudogenes were detected on 61% and 91% of nr90 chromosomes in *Aegagropila* and *Chaetomorpha*, respectively. In addition, in *Aegagropila*, some chromosomes have long ORFs (some >600 aa) with full-length transcripts aligned but no detectable homology, indicating that unknown genes are present and expressed. Heterogeneity is also extensive in *Boodlea*, whose plastome assembly contains many gene fragments, and the authors report a "large and heterogeneous population" of reads of low molecular weight without genes encoded.[14]

RNA editing of coding regions can be a feature of organellar genomes.[23,24] To inspect for RNA editing, we mapped assembled transcripts to nr90 chromosomes. For a majority of our representative genes, we found long transcripts aligned with 100% nucleotide identity across the entire coding sequence (Table S1). We also mapped both RNA and DNA Illumina reads to chromosomes, finding 402 and 332 variants exclusively in RNA reads in *Aegagropila* and *Chaetomorpha*, respectively. These can be indicative of RNA editing; however, 18,445 and 4,825 variants were exclusive to DNA reads, which greatly complicates accurate detection of editing events and also confirms that a larger population exists than what is captured in our long-read assemblies. As transcripts were found identical to genes, RNA editing of coding regions is unlikely, as was inferred in *Boodlea*.[14] It is reported that post-transcriptional addition of poly(U) or poly(A) tracts to mRNA 3′ termini is present in Cladophorales,[25] which were found in some of our transcripts immediately downstream of stop codons in both species. In other heteroplasmic organellar genomes, different gene variants have been found to be expressed without editing.[8] We found some evidence of gene variant expression in small transcript fragments, and a small number of variants were identified in both RNA and DNA Illumina mappings; however, RNA read coverage supporting these inferences was very low. Additionally, alternative transcript isoforms were not assembled for most plastid genes. Our data indicate that Cladophorales generally express a single copy per gene at high levels, but future single-molecule RNA sequencing is warranted to fully decipher gene variant expression dynamics.

Non-coding regions of minicircular chromosomes were also very dissimilar from each other across the entire plastome population. In *Aegagropila*, few chromosomes cluster at higher nucleotide identity thresholds, whereas in *Chaetomorpha*, more clusters are present (Figure 5A). There are, however, conserved motifs detected in each genome, as was found in the *Boodlea* plastome, which were found in minicircular genomes as recurring groups of motifs (Figure 5B). No motifs were found shared between species. Compared with *Aegagropila*, *Chaetomorpha* has more motifs shared across more chromosomes, and motifs are generally more conserved, including a region present on every chromosome with mean pairwise identities >90%. GC content of motif occurrences ranged from 18% to 69% in *Aegagropila* and 34% to 84% in *Chaetomorpha*. Motifs were not found in a consistent placement in respect to genes; however, in *Aegagropila*, one motif was almost always found once per chromosome, while in *Chaetomorpha*, eight motifs were almost always found once per chromosome and/or always found in a consistent orientation in respect to representative genes, which more strongly suggests they play a functional role such as transcription or replication control, while others could be a result of recombination

(B) Estimated rates of synonymous (dS) and non-synonymous (dN) mutations of 17 plastid genes, contrasting the higher rates of Cladophorales compared with those of other green algae, with larger ratios (ω) indicating relaxation of purifying selection.
(C) GC content of 17 plastid coding sequences from different algal groups, showing that later-diverging lineages have higher GC content than earlier-diverging clades. Boxplots display median and interquartile length ranges.

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

CellPress
OPEN ACCESS

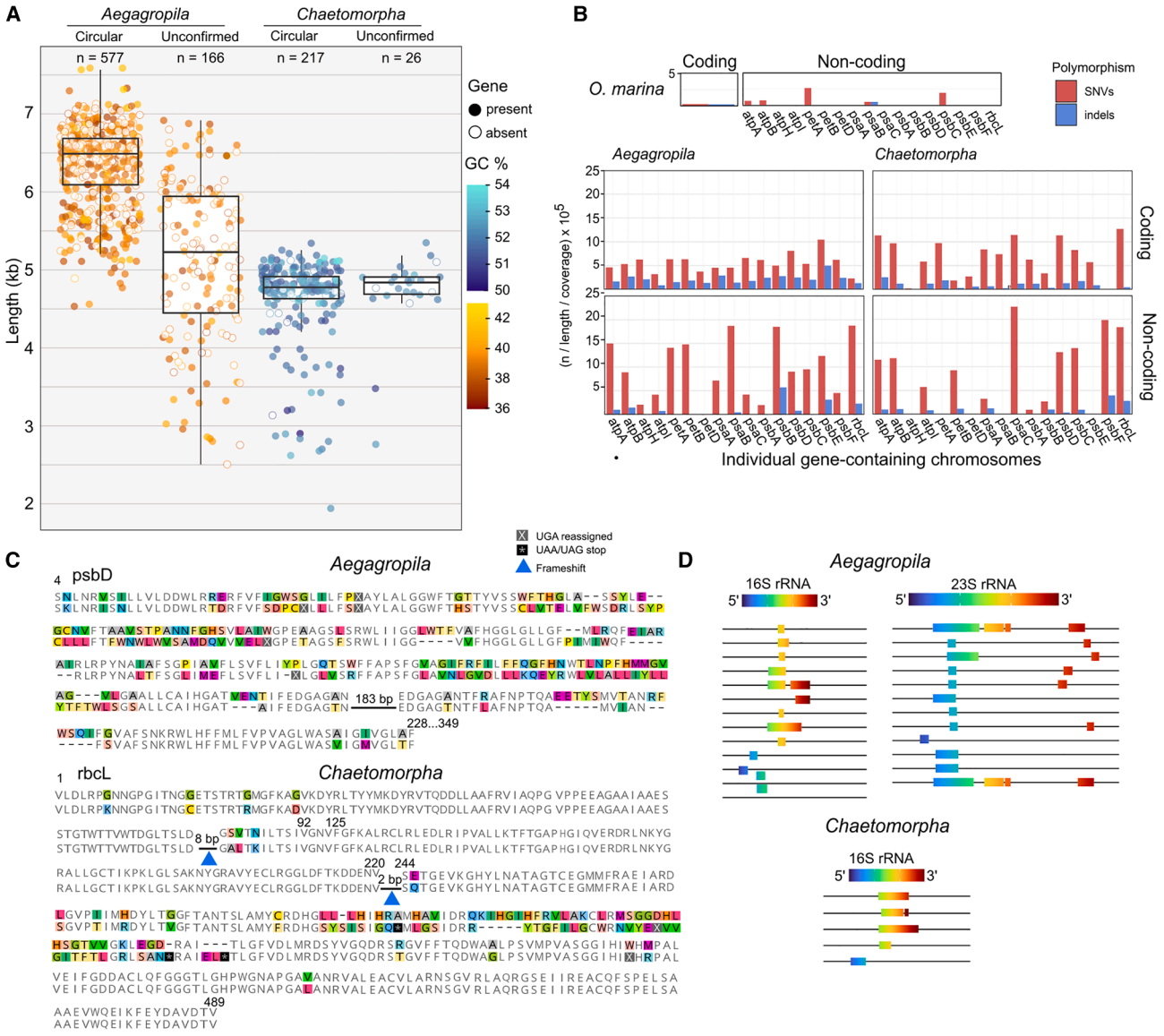

Figure 4. Minicircular chromosomes are extremely heterogeneous

(A) Length distributions of nr90 chromosomes in *Aegagropila* and *Chaetomorpha*, grouped by those that assembled as a closed circle and those that are unconfirmed in topology. Boxplots display median and interquartile length ranges. Color indicates GC content, and a filled circle indicates a gene or gene fragment is detected. See also Figure S2, Data S1A, and Table S2.

(B) Number of polymorphisms called by Illumina read alignments in coding and non-coding regions of the *O. marina* plastid (top) and individual chromosomes containing genes labeled on the *x* axis from *Aegagropila* and *Chaetomorpha* (bottom), normalized by region length and coverage.

(C) Examples of pseudogenes present on minicircular chromosomes. Amino acid sequences of best representative gene copies (top sequence) are aligned to pseudogene translations (bottom sequence). Internal reassigned UGA stop codons are in gray boxes, while UAG and UAA stop codons are in black and denoted by red triangles, and frame-shift mutations by blue triangles. Numbers above alignments indicate positions on the functional gene copy. See also Figure S5.

(D) Unique rRNA gene fragments on chromosomes of *Aegagropila* and *Chaetomorpha*. Scales above represent lengths of Ulvophyceae rRNA infernal covariance models, and colors indicate alignment regions.

between chromosomes. We did not recover any head-to-tail homoconcatemer reads that would support rolling-circle replication, nor did we recover long (i.e., longer than chromosomes) heteroconcatemer reads supporting recent recombination between different assembled chromosomes. However, we did recover reads likely resulting from recombination between the same (or very similar) chromosomes, resulting in a head-to-head configuration (Figure 5C). Two such reads from *O. marina* containing most of the plastome in a head-to-head configuration were also found, but their mechanistic origin could not be discerned.

Differential stop codon reassignment in Cladophorales lineages

In Cladophorales plastids, the UGA stop codon has been reassigned, while canonical stop codons UAA and UAG terminate

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

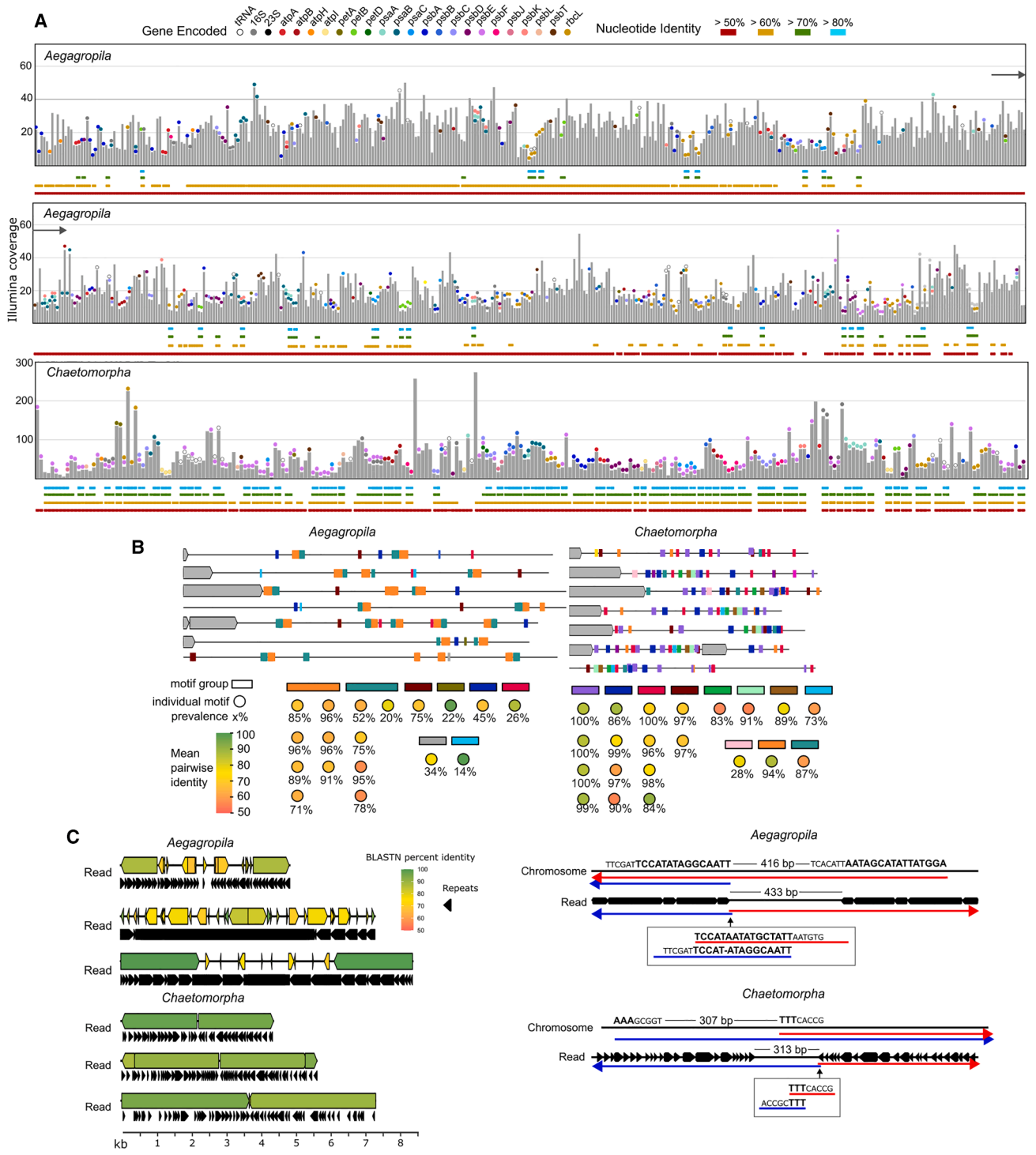

Figure 5. Analysis of sequence similarity across minicircular plastomes and evidence of recombination

(A) Illumina coverage of nr90 chromosomes, normalized by the total number of reads aligned to all chromosomes in each species. Dots display genes or gene fragments detected on each chromosome. Colored lines below plots show groups of chromosomes when clustered at various sequence identity thresholds. See also Data S1A and Table S2.

(B) Conserved motifs in *Aegagropila* and *Chaetomorpha* plastomes. Shorter motifs were found recurring in groups, and groups are displayed as colored rectangles. Circles below rectangles indicate original motifs within each group, colored by pairwise identity across all instances on all chromosomes. Percentages indicate the percentage of nr90 chromosomes each motif was found on. Gray boxes on chromosomes indicate genes.

(C) Left: examples of individual plastid-mapping IR reads from *Aegagropila* (top, PacBio HiFi) and *Chaetomorpha* (bottom, uncorrected nanopore). Colored boxes denote regions of BLASTn mappings of plastid chromosomes to reads, with the color scale showing percent similarity. Black regions below reads represent

(legend continued on next page)

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

translation.[14] However, assessment of UGA positions in the *Boodlea* was unable to determine what aa UGA now encodes.[14] Using our new plastomes, a comparison of internal UGA codons within genes shows stark differences between Okellyaceae genes and those of other clades. By examining homologous positions in other Chlorophyta, *Okellya* appears to use UGA to encode tryptophan (W), which requires a single base-pair change from UGG (Figures 6A–6C). Not all tryptophan codons have become UGA—31 and 44 UGG tryptophan codons remain in *Okellya marina* and *Okellya* sp. SL, respectively.

In other Cladophorales, as was found in *Boodlea*, homology comparisons give little indication as to UGA usage (Figures 6A and 6B). To account for the possibility of misalignments from using too-divergent genes, we collected all potentially functional copies of each gene from *Aegagropila* and *Chaetomorpha* and aligned each separately to the Chlorophyta gene set. Doing so revealed that UGA codons within better-aligned genes and regions more commonly occurred where a cysteine residue was conserved in other Chlorophyta (Figure 6C). Cysteine, typically encoded by UGU or UGC codons, also requires a single base-pair change to become UGA and is likely the reassignment in non-Okellyaceae lineages, although it is not impossible that a single codon could encode more than one amino acid.[26,27] Canonical cysteine codons are also found in these genes.

Despite UGA codons occurring near the ends of some of our representative genes (including *Boodlea* genes), in these cases BLAST alignment to other algal genes continued past these positions, and therefore, in contrast to what was reported in *Boodlea*,[14] we did not find evidence of UGA being used as a genuine stop codon, and instead, UAA or UAG were used. In addition, while AUG is likely used as a start codon in some plastid genes, in many cases the start codon was ambiguous. In all genomes, the tRNA-M(cat) identified is an elongator tRNA rather than an initiator.

The sister order of the Cladophorales is the Trentepohliales, whose chloroplast genomes have reassigned UAG and UGA stop codons to encode arginine and possibly other aa but can also be used as a genuine termination codon.[16] Analysis of Cladophorales UGA reassignments therefore shows that UGA reassignment appears to have occurred independently in Trentepohliales, Okellyaceae, and non-Okellyaceae Cladophorales.

DISCUSSION

Sequencing new Cladophorales plastid genomes has uncovered key transitions underlying their unique evolutionary trajectory. We show that Cladophorales plastid genomes were ancestrally circular and that elevated mutation rates and reduction of coding content to the current repertoire preceded fragmentation. Plastome fragmentation produced many heterogeneous minicircular chromosomes, which were an evolutionary intermediate to hairpins of *Boodlea*, and only the *Boodlea* plastome contains extensive IRs enabling hairpin formation. Comparisons of UGA codons also reveal their likely reassignment, and this reassignment occurred independently in Okellyaceae and non-Okellyaceae lineages and in sister order Trentepohliales. Figure 7 presents an outline of our findings, which allow us to formulate more informed hypotheses, particularly around genome fragmentation, gene loss and sequence divergence, and the emergence of IRs in this lineage.

An elevated mutation rate is ancestral to Cladophorales plastids, and elevated rates (though somewhat lower) are also found in plastids of the sister order Trentepohliales, whose genomes are conversely expanded and have decreased in GC content.[16] By comparison of branch lengths (Figure 3A), it appears plastid mutation rates increased in the ancestor of Cladophorales and Trentepohliales, then accelerated in Cladophorales, particularly in Okellyaceae. Mutational burden likely favored loss of non-essential genes, while other genes were rescued by gene transfer to the nucleus, as nuclear evidence of other plastid genes was found in *Boodlea*.[14] This aligns with the "mutational burden hypothesis," which proposes that the risk of deleterious mutations favors more compact genomes,[28] and can also explain non-coding sequence contraction in the Okellyaceae. However, this theory conflicts with the expansion of intergenic space of Trentepohliales plastomes. In fact, there are multiple examples where mutational load does not correlate with organelle genome size,[29] likely due to differences in the interplay of a range of both adaptive and non-adaptive processes. Terrestrial algae such as the Trentepohliales have larger plastomes in general,[16] suggesting adaptations to environmental pressures are involved in that lineage. In Cladophorales plastomes, we believe that relaxation of selection on mutations is a stronger driver of structural genomic changes.

The 21 protein-coding genes remaining in Cladophorales plastids are all major components of electron transport chain complexes, with the sole exception of *rbcL*, which encodes the large subunit of the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) complex. Twelve of these genes (*atpA*, *atpB*, *petB*, *petD*, *psaA*, *psaB*, *psbA-E*, and *psbT*) overlap with the 13 protein-coding genes retained in peridinin dinoflagellate plastids.[30] These genes are highly conserved across plastids, although individual losses of each have occurred in more gene-rich plastomes.[31] Conservation of major electron transfer components parallels gene retention patterns of reduced mitogenomes, and complementary hypotheses have been put forward to explain this phenomenon. Together, they suggest that these components must be retained to control early assembly steps of these complexes and to be subsequently regulated by complex levels (co-location for control of assembly [COCOA][30] and control by epistasy of synthesis [CES] hypotheses,[32] which can both necessitate *rbcL* retention for RuBisCO[33]) and to retain tight control of redox levels within organelles (co-location for redox regulation [CORR] hypothesis).[34]

repeats (≥20 bp, ≤30% mismatches). Right: examples of evidence that such reads result from recombination of chromosomes. Head-to-head mappings of a chromosome (i.e., plus and negative strand mappings of the same region) are shown as blue and red arrows. On reads, mappings overlap at short sequences (larger, bold text) that correspond to either palindromic sequences or separate regions on chromosomes that are identical or nearly identical in reverse complement. Therefore, recombination with plus and negative strands from original chromosomes would result in the head-to-head concatemers found.
See also Data S1C.

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

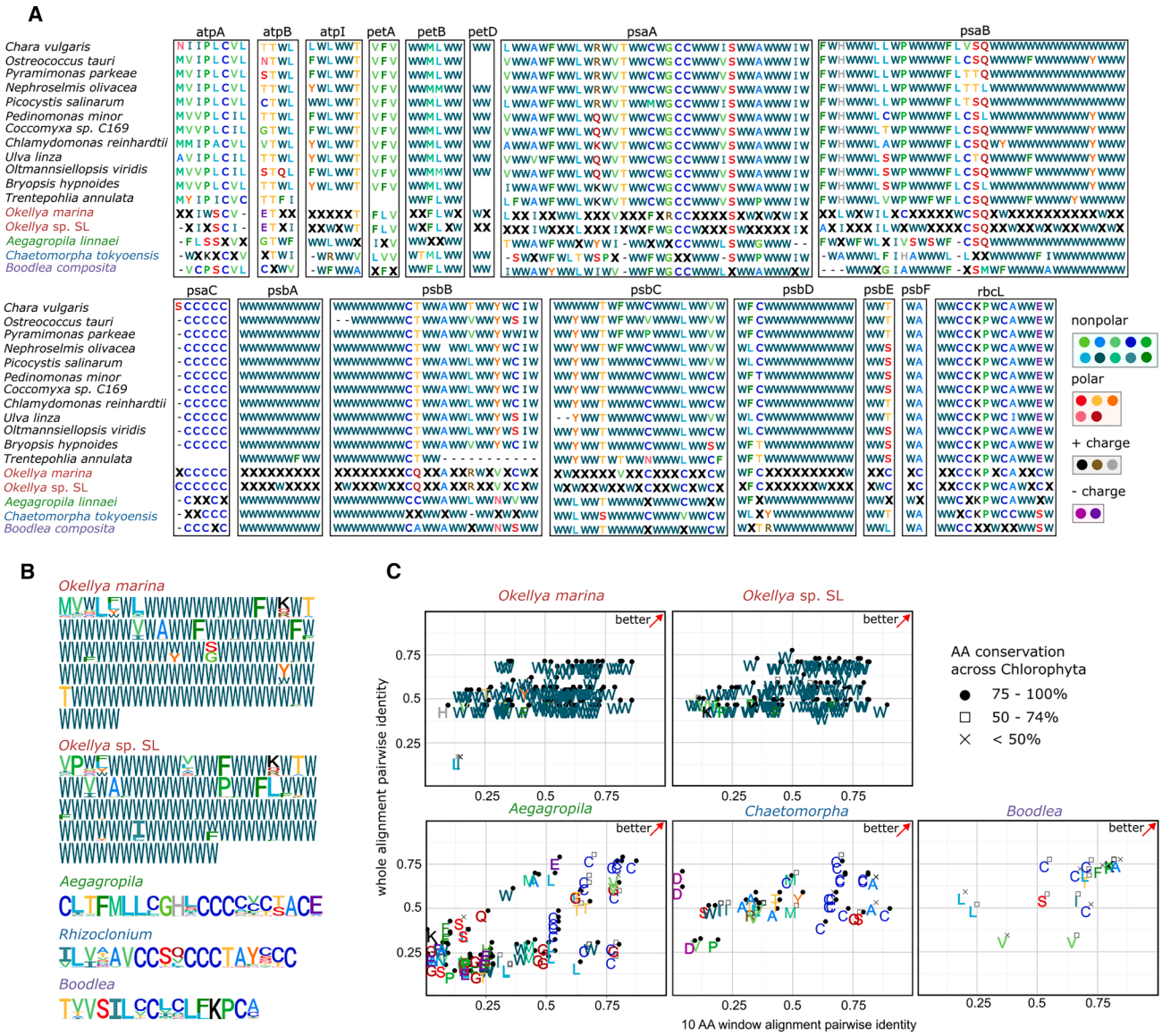

Figure 6. Concurrent stop codon reassignment in Cladophorales plastid genomes

(A) Multiple sequence alignment of Chlorophyta plastid genes, subsetted to positions containing internal UGA codons in Cladophorales, denoted by a black X.

(B) Sequence logo displaying the frequency of aa across Chlorophyta found in each UGA position, derived from the matrix in (A).

(C) Plots of aa frequencies across Chlorophyta for all UGA positions in species. Mean pairwise identity of the entire alignment of each protein to all other Chlorophyta is on the *y* axis, while mean pairwise identity of a 10 aa window centered on the UGA position is on the *x* axis. UGA positions in better-aligned proteins and regions are closer to the top right corner. Each aa shows the most common one found across Chlorophyta, and shapes denote the frequency with which it is found.

See also Data S1D and Table S1.

The fact that, over several hundred years of evolution under considerable mutation rates, Cladophorales plastomes have retained nine more protein-coding genes than those of the dinoflagellates suggests lineage-specific requirements for a minimal photosynthetic plastome. This could be due to differences in complex assembly or expression systems within the plastid, and Cladophorales plastome coding requirements can still be governed by the aforementioned hypotheses. For instance, translational feedback loops that govern the CES hypothesis are not completely preserved across green algae and land plants,[35] and therefore retention of specific genes within Cladophorales plastids may be required for proper complex regulation by the plastid.

In addition to reduced coding capacity, other evolutionary parallels are now revealed between the minicircular plastomes of peridinin dinoflagellates and those of Cladophorales. Dinoflagellate minicircular chromosomes are ~1.5–6 kb in length,[36,37] similar in size to those of *Chaetomorpha* but smaller than *Aegagropila*, with some larger minicircles up to 10 kb found.[38] They contain 1–3 genes each, which are also highly divergent in sequence,[39]

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

CellPress
OPEN ACCESS

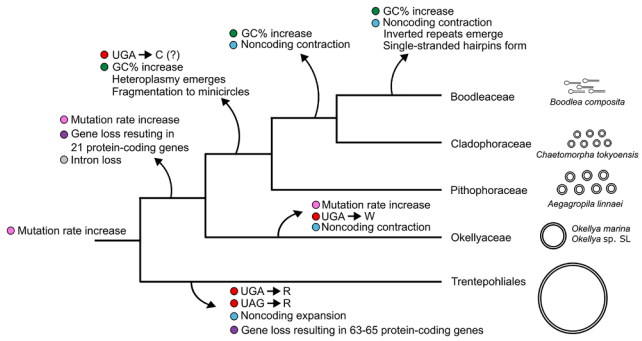

Figure 7. Overview of the evolution of Cladophorales plastid genomes and sister order Trentepohliales for comparison
Colored dots indicate evolutionary events of shared themes.

including the SSU and LSU rRNA genes, which are likely fragmented, and only a small set of tRNAs have been identified.[37,40] Interestingly, 3′-polyuridylylation editing of transcripts is common only to plastomes of both lineages,[41] as well as some species of dinoflagellate relatives Chromera.[42] GC content of dinoflagellate plastomes are also slightly elevated, but it is possible this occurred before plastome fragmentation.[39] The most striking difference between the two lineages with fragmented plastomes is the extreme heterogeneity exhibited by Cladophorales. In peridinin dinoflagellates, only a few species contain additional variants of a few genes,[43] and the number of "empty" minicircles found ranges from only 1 to 10 per species.[36,37] They also have non-coding "core" regions that are conserved within a species,[44] whereas Cladophorales non-coding regions have only small, irregular motifs. Dinoflagellate plastomes also undergo transcript editing,[45] which is not detected in Cladophorales.

Questions remain as to how mutation rates have elevated considerably in Cladophorales plastids, how heteroplasmy and fragmentation have occurred in non-Okellyaceae lineages, and how these features are connected. Elevated mutation rates and recombinational errors in several plastid genomes have been shown (experimentally or by correlation) to be associated with defects in DNA repair proteins such as MSH1,[46] RECA,[47] and WHY1/3.[48,49] Defects can also coincide with the formation of minicircles in mitochondria. For example, loss of RAD52 coincides with minicircles in Stylonematophyceae,[7] and loss of mtSSB coincides with minicircles in lice.[6] Furthermore, gene conversion (allele mismatch repair during homologous recombination) is believed to be a leading contributor to organellar homoplasmy,[50] and although we find evidence of recombination in minicircular chromosomes, gene conversion may be defective in heteroplasmic lineages. In Cladophorales, as fragmentation occurred after the divergence of Okellyaceae, if alterations in DNA repair are responsible for elevated mutation rates, heteroplasmy, and fragmentation into minicircles, there would presumably be separate mechanisms at play, one ancestral to all Cladophorales and one specific to non-Okellyaceae lineages.

Fragmentation of organellar genomes is often proposed to occur through structural errors that arise during defective DNA recombination or repair. However, an alternate theory has suggested that if heteroplasmy was able to be sustained in the ancestral intact plastome, this may have allowed multi-gene molecules to whittle down to single-gene fragments over time through the ability to withstand knockout mutations on alternate fragments.[6] For instance, if there are two copies of a chromosome, each with genes A and B, alternate knockout mutations in A and B on each chromosome copy would necessitate selection to maintain both copies in the population, each now encoding a single functional gene. The fact that fragmented plastomes are massively heterogeneous while Okellyaceae plastomes are not would be in line with this theory. In addition to dinoflagellate plastomes, the presence of non-functional gene copies and fragments has been reported in other cases of fragmented mitochondrial genomes,[51] as has the presence of empty chromosomes.[52,53] However, the extent of heterogeneity in Cladophorales appears to be unparalleled across all organelle genomes sequenced to date, and why non-functional chromosomes are allowed to persist in the population is unclear, as their replication would subject the plastome to additional energy costs. As gene conversion can also govern effective germline variant sorting,[54,55] it is possible that the proposed lack of gene conversion also results in a breakdown of germline selection against non-functional chromosomes.

Another theory to explain plastome fragmentation in *Boodlea* was the possibility of foreign DNA insertion into the plastome. It was reported that non-coding sequences of *Boodlea* plastid chromosomes were found in nuclear reads containing retrotransposons, prompting the suggestion that the introduction of repetitive sequences into the plastid from nuclear retrotransposons leads to fragmentation and hairpin topologies.[14] Like what was found in *Boodlea*, we also found plastome sequences adjacent to retrotransposon-like gene remnants in long reads from *O. marina*, *Aegagropila*, and *Chaetomorpha* (Figure S6). Given the length and GC content of reads, and at times the presence of non-plastid genes, these reads are likely nuclear. As they are present in *O. marina*, if such reads derive from retrotransposon-plastid interactions, these interactions preceded fragmentation. It is still possible that the effects of retrotransposon insertion were exacerbated by a DNA-repair-defective background specific to non-Okellyaceae lineages, precipitating fragmentation. However, plastid DNA sequences often insert into the nucleus, where they have been shown to then be associated with retrotransposons.[56–58] Insertion of mobile DNA elements into plastids is rare compared with mitochondria, likely because plastids are more resistant to DNA uptake than mitochondria,[59] but such events have been reported.[60,61] Moreover, repetitive sequences are not prevalent in minicircular plastomes. Overall, with current data, it appears more likely that nuclear plastid-derived sequences in Cladophorales result from DNA transfer out of the plastid rather than the introduction of those nuclear sequences into the plastid.

If repetitive sequences of *Boodlea* do not derive from retrotransposons, the transition from IR-lacking minicircles to single-stranded hairpins is drastic, particularly if gene conversion is defective in these plastomes, as this is proposed to create homogenous telomeric-type and other repeat sequences in linear genomes.[62] However, we report reads with head-to-head configurations (i.e., IRs) in Cladophorales that likely formed via recombination (Figure 5C). In *Aegagropila* and *Chaetomorpha*, 51 and 96 such reads were found. Their presence may signify a pathway for *Boodlea*-type chromosomes to form. Theoretically, *Boodlea* plastome replication mechanisms would have needed to be ancestrally established to allow the emergence of hairpin

chromosomes, and this replication system may have been able to replicate recombination artifacts such as these. Currently, replication mechanisms of Cladophorales hairpins are unknown, as well as how hairpin ends are protected from exonuclease activity. Replication in *Boodlea* is possibly not hairpin-dependent, as not all assembled linear chromosomes appear to form hairpins, and in other *Siphonocladus* clade members, no evidence of terminally bound proteins that might assist in replication of linear sequences was found,[63] although other experiments suggest ends are protected in some manner.[17] It is currently unclear if there are evolutionary advantages to hairpin formation in the absence of gene conversion, but one possibility is that replication systems involved with linear or hairpin chromosomes are more efficient, leading to their propagation. Additionally, drastic mutations in hairpin chromosomes would more likely lead to loss of structural integrity, leading to better selection against these.

Stronger base pairing resulting from high GC content of the *Siphonocladus* clade is likely a structural necessity for stable hairpin chromosomes. Plastome GC content increases in a phylogenetic manner across Cladophorales lineages, similar to minicircular mitogenomes of Stylonematophyceae.[7] Changes in GC content can occur for a variety of reasons. One common process is DNA repair by GC-biased gene conversion, which appears to be defective in Cladophorales, although other DNA repair mechanisms can be GC-biased.[64] Environmental factors can select for GC changes; however, species here are sampled from similar environments. In Cladophorales, GC increase appears to be linked to fragmentation and has an inverse relationship with mean chromosome size, but ultimately, the drivers behind GC changes are unknown.

Overall, Cladophorales plastid genomes present an extraordinary case study on genome evolution, and many questions remain that warrant further investigation of this lineage. For instance, how do plastid functions remain unimpaired despite streamlining and extreme sequence divergence in core components? How are highly fragmented 16S and 23S genes functional? Why do hundreds of non-functional chromosomes persist in fragmented plastomes? How does replication occur on vastly different topologies? Furthermore, sequencing of families within the *Siphonocladus* clade more closely related to *Boodlea* is warranted, as this would greatly assist in understanding the emergence of IRs that characterize this unique plastome.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Saelin Bjornson (saelinbjornson@gmail.com).

Materials availability

This study did not generate any new materials or reagents. Recipes for modified culture media or for buffers used in DNA extraction are directly given in the STAR Methods and referenced in the key resources table. *Okellya marina* strain 56A has been deposited in the Australian National Algal Culture Collection under access number CS-1486. Further inquiries on the availability of algae cultures used in the study can be directed to the lead contact.

Data and code availability

- Raw sequence data have been deposited in the NCBI Sequence Read Archive under BioProject: PRJNA1331326, with accession numbers listed in the key resources table.
- All assembled sequence data are publicly available as of the date of publication at Mendeley Data: https://doi.org/10.17632/684wwdffjh.1.
- Graphical workflows of selected data analysis methods are provided in the supplemental information as Data S1A–S1D.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

Funding was provided by the Fundação para a Ciência e a Tecnologia (https://doi.org/10.54499/2023.06155.CEECIND/CP2845/CT0004 to H.V.), the Australian Research Council (DP200101613 to H.V.), the Australian Biological Resources Study (activity 4-G046WSD to H.V.), and the National Health and Medical Research Council (GNT1196841 to M.B.C.). O.D.C. is indebted to the European Marine Biological Resource Centre Belgium, funded by the Research Foundation – Flanders (project I001625N). J.L.S. is a Howard Hughes Medical Institute Awardee of the Life Sciences Research Foundation. This research was also supported by the Commonwealth through an Australian Government Research Training Program Scholarship (https://doi.org/10.82133/C42F-K220) as well as the University of Melbourne's Research Computing Services and the Petascale Campus Initiative.

AUTHOR CONTRIBUTIONS

H.V. and S.B. conceptualized the study. J.W. and R.W. provided algal cultures. R.H., O.M., R.D.P.-I., C.M.C.C., M.B.C., and T.J. assisted with DNA and RNA extraction and sequencing. T.T.B. assembled the *O. marina* plastid genome. O.D.C. and F.L. sequenced and assembled the *Okellya* sp. SL plastid genome. K.U. and S.D. assisted in phylogeny and dN/dS analysis. S.B. carried out sequence analysis and generated figures. H.V. and J.L.S. supervised the study. S.B. wrote the first draft of the manuscript, and all authors helped revise the manuscript, with particular input on the main text from O.D.C., R.D.P.-I., F.L., J.L.S., and H.V.

DECLARATION OF INTERESTS

R.D.P.-I. and M.B.C. have received financial support from Oxford Nanopore Technologies (ONT) to present their findings at scientific conferences. ONT played no role in study design, execution, analysis, or publication. J.L.S. is an advisor to ForensisGroup Inc. and a scientific consultant to FutureHouse Inc. For the duration of the project, J.L.S. was a Bioinformatics Visiting Scholar at MantleBio Inc.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Algae collection and culturing
- METHOD DETAILS
  - DNA and RNA extraction and sequencing
  - Plastid genome assembly
  - Delimitation of final *Aegagropila* and *Chaetomorpha* chromosomes and confirming circularity
  - Annotation of plastid genomes and selection of representative genes
  - 18S phylogeny for species confirmation
  - Multi-gene phylogeny
  - Estimates of dN/dS in algal genes
  - Polymorphism quantification on individual chromosomes
  - Illumina coverage estimation of *Aegagropila* and *Chaetomorpha* chromosomes and iterative sequence clustering
  - Transcriptome assembly, identification of gene variant expression and RNA editing
  - Motif identification across minicircular chromosomes

- Estimation of internal UGA reassignments
- Search for trans-splicing of fragmented rRNA genes
- Search for homoconcatemers and heteroconcatemers in reads
- Identification of plastid-mapping reads with inverted structure
- Identification of reads with plastid sequences and retrotransposon-type elements

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2026.02.038.

Received: December 15, 2025
Revised: January 12, 2026
Accepted: February 18, 2026

REFERENCES

1. Smith, D.R., Hua, J., and Lee, R.W. (2010). Evolution of linear mitochondrial DNA in three known lineages of Polytomella. Curr. Genet. *56*, 427–438. https://doi.org/10.1007/s00294-010-0311-5.
2. Alverson, A.J., Rice, D.W., Dickinson, S., Barry, K., and Palmer, J.D. (2011). Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. Plant Cell *23*, 2499–2513. https://doi.org/10.1105/tpc.111.087189.
3. Cheng, W.X., Wang, J., Mao, M.L., Lu, Y.B., and Zou, J.X. (2024). The mitochondrial genome of Bottapotamon fukienense (Brachiura: Potamidae) is fragmented into two chromosomes. BMC Genomics *25*, 755. https://doi.org/10.1186/s12864-024-10657-9.
4. Nadimi, M., Stefani, F.O.P., and Hijri, M. (2014). The mitochondrial genome of the Glomeromycete Rhizophagus sp. DAOM 213198 reveals an unusual organization consisting of two circular chromosomes. Genome Biol. Evol. *7*, 96–105. https://doi.org/10.1093/gbe/evu268.
5. Stampar, S.N., Broe, M.B., Macrander, J., Reitzel, A.M., Brugler, M.R., and Daly, M. (2019). Linear mitochondrial genome in Anthozoa (Cnidaria): a case study in Ceriantharia. Sci. Rep. *9*, 6094. https://doi.org/10.1038/s41598-019-42621-z.
6. Sweet, A.D., Johnson, K.P., Cameron, S.L., Sweet, A.D., Johnson, K.P., and Cameron, S.L. (2022). Independent evolution of highly variable, fragmented mitogenomes of parasitic lice. Commun. Biol. *5*, 677. https://doi.org/10.1038/s42003-022-03625-0.
7. Lee, Y., Cho, C.H., Noh, C., Yang, J.H., Park, S.I., Lee, Y.M., West, J.A., Bhattacharya, D., Jo, K., and Yoon, H.S. (2023). Origin of minicircular mitochondrial genomes in red algae. Nat. Commun. *14*, 3363. https://doi.org/10.1038/s41467-023-39084-2.
8. Yu, R., Sun, C., Zhong, Y., Liu, Y., Sanchez-Puerta, M.V., Mower, J.P., and Zhou, R. (2022). The minicircular and extremely heteroplasmic mitogenome of the holoparasitic plant *Rhopalocnemis phalloides*. Curr. Biol. *32*, 470–479.e5. https://doi.org/10.1016/j.cub.2021.11.053.
9. Spencer, D.F., and Gray, M.W. (2011). Ribosomal RNA genes in *Euglena gracilis* mitochondrial DNA: fragmented genes in a seemingly fragmented genome. Mol. Genet. Genomics *285*, 19–31. https://doi.org/10.1007/s00438-010-0585-9.
10. Lavrov, D.V., and Lang, B.F. (2014). Mitochondrial genomes in unicellular relatives of animals. In Molecular Life Sciences, R.D. Wells, J.S. Bond, J. Klinman, and B.S.S. Masters, eds. (Springer), pp. 1–4. https://doi.org/10.1007/978-1-4614-6436-5_178-2.
11. Wu, Z., Cuthbert, J.M., Taylor, D.R., and Sloan, D.B. (2015). The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. Proc. Natl. Acad. Sci. USA *112*, 10185–10191. https://doi.org/10.1073/pnas.1421397112.
12. Watanabe, S., Fučíková, K., Lewis, L.A., and Lewis, P.O. (2016). Hiding in plain sight: Koshicola spirodelophila gen. et sp. nov. (Chaetopeltidales, Chlorophyceae), a novel green alga associated with the aquatic angiosperm *Spirodela polyrhiza*. Am. J. Bot. *103*, 865–875. https://doi.org/10.3732/ajb.1500481.
13. Tang, L., Tam, N.F.Y., Lam, W., Lee, T.C.H., Xu, S.J.L., Lee, C.L., and Lee, F.W.F. (2024). Interpreting the complexities of the plastid genome in dinoflagellates: a mini-review of recent advances. Plant Mol. Biol. *114*, 114. https://doi.org/10.1007/s11103-024-01511-3.
14. Del Cortona, A., Leliaert, F., Bogaert, K.A., Turmel, M., Boedeker, C., Janouškovec, J., Lopez-Bautista, J.M., Verbruggen, H., Vandepoele, K., and De Clerck, O.D. (2017). The plastid genome in cladophorales green algae is encoded by hairpin chromosomes. Curr. Biol. *27*, 3771–3782.e6. https://doi.org/10.1016/j.cub.2017.11.004.
15. Yang, T., Sahu, S.K., Yang, L., Liu, Y., Mu, W., Liu, X., Strube, M.L., Liu, H., and Zhong, B. (2022). Comparative analyses of 3,654 plastid genomes unravel insights into evolutionary dynamics and phylogenetic discordance of green plants. Front. Plant Sci. *13*, 808156. https://doi.org/10.3389/fpls.2022.808156.
16. Fang, J., Chen, Y., Liu, G., Verbruggen, H., and Zhu, H. (2021). Chloroplast genome traits correlate with organismal complexity and ecological traits in Chlorophyta. Front. Ecol. Evol. *9*, 791166. https://doi.org/10.3389/fevo.2021.791166.
17. La Claire, J.W., Zuccarello, G.C., and Tong, S. (1997). Abundant plasmid-like DNA in various members of the orders siphonocladales and cladophorales (Chlorophyta). J. Phycol. *33*, 830–837. https://doi.org/10.1111/j.0022-3646.1997.00830.x.
18. Hou, Z., Ma, X., Shi, X., Li, X., Yang, L., Xiao, S., De Clerck, O., Leliaert, F., and Zhong, B. (2022). Phylotranscriptomic insights into a Mesoproterozoic–Neoproterozoic origin and early radiation of green seaweeds (Ulvophyceae). Nat. Commun. *13*, 1610. https://doi.org/10.1038/s41467-022-29282-9.
19. Boedeker, C., O'Kelly, C.J., Star, W., and Leliaert, F. (2012). Molecular phylogeny and taxonomy of the *Aegagropila* clade (Cladophorales, Ulvophyceae), including the description of *Aegagropilopsis* Gen Nov. and *Pseudocladophora*. J. Phycol. *48*, 808–825. https://doi.org/10.1111/j.1529-8817.2012.01145.x.
20. Wetherbee, R., Bringloe, T.T., Leliaert, F., and Verbruggen, H. (2024). Molecular data and culture observations show that the microfilamentous marine alga *Uronema marinum* Womersley is a member of the genus *Okellya* Leliaert & Rueness (Cladophorales, Chlorophyta). Cryptogam. Algol. *45*, 31–37. https://doi.org/10.5252/cryptogamiealgologie2024v45a3.
21. Xiang, Q.P., Tang, J.Y., Yu, J.G., Smith, D.R., Zhu, Y.M., Wang, Y.R., Kang, J.S., Yang, J., and Zhang, X.C. (2022). The evolution of extremely diverged plastomes in Selaginellaceae (lycophyte) is driven by repeat patterns and the underlying DNA maintenance machinery. Plant J. *111*, 768–784. https://doi.org/10.1111/tpj.15851.
22. Oldenburg, D.J., and Bendich, A.J. (2004). Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. J. Mol. Biol. *335*, 953–970. https://doi.org/10.1016/j.jmb.2003.11.020.
23. Kaur, B., Záhonová, K., Valach, M., Faktorová, D., Prokopchuk, G., Burger, G., and Lukeš, J. (2020). Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplonemids. Nucleic Acids Res. *48*, 2694–2708. https://doi.org/10.1093/nar/gkz1215.
24. Oldenkott, B., Yamaguchi, K., Tsuji-Tsukinoki, S., Knie, N., and Knoop, V. (2014). Chloroplast RNA editing going extreme: more than 3400 events of C-to-U editing in the chloroplast transcriptome of the lycophyte *Selaginella uncinata*. RNA *20*, 1499–1506. https://doi.org/10.1261/rna.045575.114.
25. Meade, M.J., Proulex, G.C.R., Manoylov, K.M., and Cahoon, A.B. (2020). Chloroplast mRNAs are 3′ polyuridylylated in the green alga *Pithophora roettleri* (Cladophorales). J. Phycol. *56*, 1124–1134. https://doi.org/10.1111/jpy.13033.

26. Suzuki, T., Ueda, T., and Watanabe, K. (1997). The 'polysemous' codon—a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. EMBO J. *16*, 1122–1134. https://doi.org/10.1093/emboj/16.5.1122.

27. Turanov, A.A., Lobanov, A.V., Fomenko, D.E., Morrison, H.G., Sogin, M.L., Klobutcher, L.A., Hatfield, D.L., and Gladyshev, V.N. (2009). Genetic code supports targeted insertion of two amino acids by one codon. Science *323*, 259–261. https://doi.org/10.1126/science.1164748.

28. Lynch, M., Koskella, B., and Schaack, S. (2006). Mutation pressure and the evolution of organelle genomic architecture. Science *311*, 1727–1730. https://doi.org/10.1126/science.1118884.

29. Smith, D.R. (2016). The mutational hazard hypothesis of organelle genome evolution: 10 years on. Mol. Ecol. *25*, 3769–3775. https://doi.org/10.1111/mec.13742.

30. Howe, C.J., and Barbrook, A.C. (2024). Dinoflagellate chloroplasts as a model for extreme genome reduction and fragmentation in organelles–the COCOA principle for gene retention. Protist *175*, 126048. https://doi.org/10.1016/j.protis.2024.126048.

31. Mohanta, T.K., Mishra, A.K., Khan, A., Hashem, A., Abd Allah, E.F., and Al-Harrasi, A. (2020). Gene loss and evolution of the plastome. Genes *11*, 1133. https://doi.org/10.3390/genes11101133.

32. Wollman, F.A., Minai, L., and Nechushtai, R. (1999). The biogenesis and assembly of photosynthetic proteins in thylakoid membranes1. Biochim. Biophys. Acta *1411*, 21–85. https://doi.org/10.1016/S0005-2728(99)00043-2.

33. Wietrzynski, W., Traverso, E., Wollman, F.A., and Wostrikoff, K. (2021). The state of oligomerization of Rubisco controls the rate of synthesis of the Rubisco large subunit in *Chlamydomonas reinhardtii*. Plant Cell *33*, 1706–1727. https://doi.org/10.1093/plcell/koab061.

34. de Paula, W.B.M., Allen, J.F., and van der Giezen, M. (2012). Mitochondria, hydrogenosomes and mitosomes in relation to the CoRR hypothesis for genome function and evolution. In Organelle Genetics, C.E. Bullerwell, ed. (Springer), pp. 105–119. https://doi.org/10.1007/978-3-642-22380-8_5.

35. Ghandour, R., Gao, Y., Ruf, S., Bock, R., and Zoschke, R. (2025). Assembly-dependent translational feedback regulation of photosynthetic proteins in land plants. Nat. Plants *11*, 1920–1938. https://doi.org/10.1038/s41477-025-02074-x.

36. He, J., Huang, Y., Li, L., Lin, S., Ma, M., Wang, Y., and Lin, S. (2024). Novel plastid genome characteristics in Fugacium Kawagutii and the trend of accelerated evolution of plastid proteins in dinoflagellates. Genome Biol. Evol. *16*, evad237. https://doi.org/10.1093/gbe/evad237.

37. Barbrook, A.C., Santucci, N., Plenderleith, L.J., Hiller, R.G., and Howe, C.J. (2006). Comparative analysis of dinoflagellate chloroplast genomes reveals rRNA and tRNA genes. BMC Genomics *7*, 297. https://doi.org/10.1186/1471-2164-7-297.

38. Nelson, M.J., and Green, B.R. (2005). Double hairpin elements and tandem repeats in the non-coding region of Adenoides eludens chloroplast gene minicircles. Gene *358*, 102–110. https://doi.org/10.1016/j.gene.2005.05.024.

39. Dorrell, R.G., Klinger, C.M., Newby, R.J., Butterfield, E.R., Richardson, E., Dacks, J.B., Howe, C.J., Nisbet, E.R., and Bowler, C. (2017). Progressive and biased divergent evolution underpins the origin and diversification of peridinin dinoflagellate plastids. Mol. Biol. Evol. *34*, 361–379. https://doi.org/10.1093/molbev/msw235.

40. Nelson, M.J., Dang, Y., Filek, E., Zhang, Z., Yu, V.W.C., Ishida, K.I., and Green, B.R. (2007). Identification and transcription of transfer RNA genes in dinoflagellate plastid minicircles. Gene *392*, 291–298. https://doi.org/10.1016/j.gene.2007.01.018.

41. Wang, Y., and Morse, D. (2006). Rampant polyuridylylation of plastid gene transcripts in the dinoflagellate Lingulodinium. Nucleic Acids Res. *34*, 613–619. https://doi.org/10.1093/nar/gkj438.

42. Dorrell, R.G., Drew, J., Nisbet, R.E.R., and Howe, C.J. (2014). Evolution of chloroplast transcript processing in Plasmodium and its chromerid algal relatives. PLOS Genet. *10*, e1004008. https://doi.org/10.1371/journal.pgen.1004008.

43. Iida, S., Kobiyama, A., Ogata, T., and Murakami, A. (2010). Differential DNA rearrangements of plastid genes, psbA and psbD, in two species of the dinoflagellate Alexandrium. Plant Cell Physiol. *51*, 1869–1877. https://doi.org/10.1093/pcp/pcq152.

44. Zhang, Z., Cavalier-Smith, T., and Green, B.R. (2002). Evolution of dinoflagellate unigenic minicircles and the partially concerted divergence of their putative replicon origins. Mol. Biol. Evol. *19*, 489–500. https://doi.org/10.1093/oxfordjournals.molbev.a004104.

45. Klinger, C.M., Paoli, L., Newby, R.J., Wang, M.Y.W., Carroll, H.D., Leblond, J.D., Howe, C.J., Dacks, J.B., Bowler, C., Cahoon, A.B., et al. (2018). Plastid transcript editing across dinoflagellate lineages shows lineage-specific application but conserved trends. Genome Biol. Evol. *10*, 1019–1038. https://doi.org/10.1093/gbe/evy057.

46. Wu, Z., Waneka, G., Broz, A.K., King, C.R., and Sloan, D.B. (2020). MSH1 is required for maintenance of the low mutation rates in plant mitochondrial and plastid genomes. Proc. Natl. Acad. Sci. USA *117*, 16448–16455. https://doi.org/10.1073/pnas.2001998117.

47. Rowan, B.A., Oldenburg, D.J., and Bendich, A.J. (2010). RecA maintains the integrity of chloroplast DNA molecules in *Arabidopsis*. J. Exp. Bot. *61*, 2575–2588. https://doi.org/10.1093/jxb/erq088.

48. Zhang, J., Ruhlman, T.A., Sabir, J.S.M., Blazier, J.C., Weng, M.L., Park, S., and Jansen, R.K. (2016). Coevolution between nuclear-encoded DNA replication, recombination, and repair genes and plastid genome complexity. Genome Biol. Evol. *8*, 622–634. https://doi.org/10.1093/gbe/evw033.

49. Maréchal, A., Parent, J.-S., Véronneau-Lafortune, F., Joyeux, A., Lang, B.F., and Brisson, N. (2009). Whirly proteins maintain plastid genome stability in Arabidopsis. Proc. Natl. Acad. Sci. USA *106*, 14693–14698. https://doi.org/10.1073/pnas.0901710106.

50. Khakhlova, O., and Bock, R. (2006). Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J. *46*, 85–94. https://doi.org/10.1111/j.1365-313X.2006.02673.x.

51. Burger, G., Forget, L., Zhu, Y., Gray, M.W., and Lang, B.F. (2003). Unique mitochondrial genome architecture in unicellular relatives of animals. Proc. Natl. Acad. Sci. USA *100*, 892–897. https://doi.org/10.1073/pnas.0336115100.

52. Sloan, D.B., Alverson, A.J., Chuckalovcak, J.P., Wu, M., McCauley, D.E., Palmer, J.D., and Taylor, D.R. (2012). Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLOS Biol. *10*, e1001241. https://doi.org/10.1371/journal.pbio.1001241.

53. Sanchez-Puerta, M.V., García, L.E., Wohlfeiler, J., and Ceriotti, L.F. (2017). Unparalleled replacement of native mitochondrial genes by foreign homologs in a holoparasitic plant. New Phytol. *214*, 376–387. https://doi.org/10.1111/nph.14361.

54. Edwards, D.M., Røyrvik, E.C., Chustecki, J.M., Giannakis, K., Glastad, R.C., Radzvilavicius, A.L., and Johnston, I.G. (2021). Avoiding organelle mutational meltdown across eukaryotes with or without a germline bottleneck. PLOS Biol. *19*, e3001153. https://doi.org/10.1371/journal.pbio.3001153.

55. Broz, A.K., Keene, A., Fernandes Gyorfy, M., Hodous, M., Johnston, I.G., and Sloan, D.B. (2022). Sorting of mitochondrial and plastid heteroplasmy in *Arabidopsis* is extremely rapid and depends on MSH1 activity. Proc. Natl. Acad. Sci. USA *119*, e2206973119. https://doi.org/10.1073/pnas.2206973119.

56. Michalovová, M., Vyskot, B., and Kejnovsky, E. (2013). Analysis of plastid and mitochondrial DNA insertions in the nucleus (NUPTs and NUMTs) of six plant species: size, relative age and chromosomal localization. Heredity *111*, 314–320. https://doi.org/10.1038/hdy.2013.51.

57. Marczuk-Rojas, J.P., Salmerón, A., Alcayde, A., Isanbaev, V., and Carretero-Paulet, L. (2024). Plastid DNA is a major source of nuclear

genome complexity and of RNA genes in the orphan crop moringa. BMC Plant Biol. *24*, 437. https://doi.org/10.1186/s12870-024-05158-6.

58. Calatrava, V., Stephens, T.G., Gabr, A., Bhaya, D., Bhattacharya, D., and Grossman, A.R. (2022). Retrotransposition facilitated the establishment of a primary plastid in the thecate amoeba *Paulinella*. Proc. Natl. Acad. Sci. USA *119*, e2121241119. https://doi.org/10.1073/pnas.2121241119.
59. Smith, D.R. (2014). Mitochondrion-to-plastid DNA transfer: it happens. New Phytol. *202*, 736–738. https://doi.org/10.1111/nph.12704.
60. Iorizzo, M., Senalik, D., Szklarczyk, M., Grzebelus, D., Spooner, D., and Simon, P. (2012). De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. BMC Plant Biol. *12*, 61. https://doi.org/10.1186/1471-2229-12-61.
61. Cremen, M.C.M., Leliaert, F., Marcelino, V.R., and Verbruggen, H. (2018). Large diversity of nonstandard genes and dynamic evolution of chloroplast genomes in siphonous green algae (Bryopsidales, Chlorophyta). Genome Biol. Evol. *10*, 1048–1061. https://doi.org/10.1093/gbe/evy063.
62. Smith, D.R., and Keeling, P.J. (2013). Gene conversion shapes linear mitochondrial genome architecture. Genome Biol. Evol. *5*, 905–912. https://doi.org/10.1093/gbe/evt059.
63. La Claire, J.W., and Wang, J. (2004). Structural characterization of the terminal domains of linear plasmid-like DNA from the green alga *Ernodesmis* (Chlorophyta). J. Phycol. *40*, 1089–1097. https://doi.org/10.1111/j.1529-8817.2004.04100.x.
64. Weissman, J.L., Fagan, W.F., and Johnson, P.L.F. (2019). Linking high GC content to the repair of double strand breaks in prokaryotic genomes. PLOS Genet. *15*, e1008493. https://doi.org/10.1371/journal.pgen.1008493.
65. Keller, M.D., Selvin, R.C., Claus, W., and Guillard, R.R.L. (1987). Media for the culture of oceanic ultraphytoplankton 1, 2. J. Phycol. *23*, 633–638. https://doi.org/10.1111/j.1529-8817.1987.tb04217.x.
66. West, J.A., and McBride, D.L. (1999). Long-term and diurnal carpospore discharge patterns in the Ceramiaceae, Rhodomelaceae and Delesseriaceae (Rhodophyta). Hydrobiologia *398–399*, 101–114. https://doi.org/10.1023/A:1017025815001.
67. Hossen, R., Courtney, M., Sim, A., Khan, M.A.A.K., Verbruggen, H., and Bringloe, T. (2025). An optimized CTAB method for genomic DNA extraction from green seaweeds (Ulvophyceae). Appl. Plant Sci. *13*, e11625. https://doi.org/10.1002/aps3.11625.
68. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
69. Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. *37*, 540–546. https://doi.org/10.1038/s41587-019-0072-8.
70. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. (2020). Using SPAdes de novo assembler. Curr. Protoc. Bioinform. *70*, e102. https://doi.org/10.1002/cpbi.102.
71. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. *27*, 722–736. https://doi.org/10.1101/gr.215087.116.
72. Crane, C.F., Nemacheck, J.A., Subramanyam, S., Williams, C.E., and Goodwin, S.B. (2022). slag: a program for seeded local assembly of genes in complex genomes. Mol. Ecol. Resour. *22*, 1999–2017. https://doi.org/10.1111/1755-0998.13580.
73. Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLOS One *11*, e0163962. https://doi.org/10.1371/journal.pone.0163962.
74. Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics *29*, 2933–2935. https://doi.org/10.1093/bioinformatics/btt509.
75. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158.
76. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.
77. Wickramarachchi, A., Mallawaarachchi, V., Rajan, V., and Lin, Y. (2020). Metabcc-lr: meta genomics binning by coverage and composition for long reads. Bioinformatics *36*, i3–i11. https://doi.org/10.1093/bioinformatics/btaa441.
78. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
79. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192. https://doi.org/10.1093/bib/bbs017.
80. Qin, M., Wu, S., Li, A., Zhao, F., Feng, H., Ding, L., and Ruan, J. (2019). LRScaf: improving draft genomes using long noisy reads. BMC Genomics *20*, 955. https://doi.org/10.1186/s12864-019-6337-2.
81. Beck, N., and Lang, B. (2019). MFannot, Organelle Genome Annotation Webserver (Université de Montréal).
82. Chan, P.P., Lin, B.Y., Mak, A.J., and Lowe, T.M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. *49*, 9077–9096. https://doi.org/10.1093/nar/gkab688.
83. Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. *9*, 286–298. https://doi.org/10.1093/bib/bbn013.
84. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
85. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589. https://doi.org/10.1038/nmeth.4285.
86. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.
87. Steenwyk, J.L., Martínez-Redondo, G.I., Buida, T.J., III, Gluck-Thaler, E., Shen, X.X., Gabaldón, T., Rokas, A., and Fernández, R. (2024). PhyKIT: a multitool for phylogenomics. Curr. Protoc. *4*, e70016. https://doi.org/10.1002/cpz1.70016.
88. Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., and Calcott, B. (2017). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. *34*, 772–773. https://doi.org/10.1093/molbev/msw260.
89. Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. *34*, W609–W612. https://doi.org/10.1093/nar/gkl315.
90. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088.
91. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.
92. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

93. Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. Preprint at arXiv. https://doi.org/10.48550/arXiv.1207.3907.

94. Quinlan, A.R. (2014). BEDTools: the Swiss-army tool for genome feature analysis. Curr. Protoc. Bioinform. *47*, 11.12.1–11.1234. https://doi.org/10.1002/0471250953.bi1112s47.

95. James, B.T., Luczak, B.B., and Girgis, H.Z. (2018). MeShClust: an intelligent tool for clustering DNA sequences. Nucleic Acids Res. *46*, e83. https://doi.org/10.1093/nar/gky315.

96. Kent, W.J. (2002). BLAT—the BLAST-like alignment tool. Genome Res. *12*, 656–664. https://doi.org/10.1101/gr.229202.

97. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.

98. Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME suite. Nucleic Acids Res. *43*, W39–W49. https://doi.org/10.1093/nar/gkv416.

99. Wilkins, D. (2023). Gggenes: Draw Gene Arrow Maps in 'ggplot2'. R Package Version 0.5.0.

100. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.

101. Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics *33*, 3645–3647. https://doi.org/10.1093/bioinformatics/btx469.

102. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580. https://doi.org/10.1093/nar/27.2.573.

103. Nakamura-Gouvea, N., Alves-Lima, C., Benites, L.F., Iha, C., Maracaja-Coutinho, V., Aliaga-Tobar, V., Araujo Amaral Carneiro, M., Yokoya, N.S., Marinho-Soriano, E., Graminha, M.A.S., et al. (2022). Insights into agar and secondary metabolite pathways from the genome of the red alga Gracilaria domingensis (Rhodophyta, Gracilariales). J. Phycol. *58*, 406–423. https://doi.org/10.1111/jpy.13238.

104. Ontiveros-Palacios, N., Cooke, E., Nawrocki, E.P., Triebel, S., Marz, M., Rivas, E., Griffiths-Jones, S., Petrov, A.I., Bateman, A., and Sweeney, B. (2025). Rfam 15: RNA families database in 2025. Nucleic Acids Res. *53*, D258–D267. https://doi.org/10.1093/nar/gkae1023.

105. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. *41*, D590–D596. https://doi.org/10.1093/nar/gks1219.

106. Boedeker, C., Wynne, M.J., and Zuccarello, G.C. (2023). Hidden diversity in high-latitude Southern Hemisphere environments: Reinstatement of the genus *Rama* and description of *Vandenhoekia* gen. nov. (Cladophoraceae, Ulvophyceae, Chlorophyta), two highly variable genera. J. Phycol. *59*, 1284–1298. https://doi.org/10.1111/jpy.13394.

107. Ichihara, K., Shimada, S., and Miyaji, K. (2013). Systematics of *Rhizoclonium*-like algae (Cladophorales, Chlorophyta) from Japanese brackish waters, based on molecular phylogenetic and morphological analyses. Phycologia *52*, 398–410. https://doi.org/10.2216/12-102.1.

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
HB buffer	This study	N/A
LB buffer	This study	N/A
K-enriched seawater medium	Keller et al.[65]	N/A
Modified Provasoli's Enriched Seawater	West and McBride[66]	N/A
CTAB lysis buffer	Hossen et al.[67]	N/A
Critical commercial assays		
RNeasy Plant Mini Kit	QIAGEN	Cat no./ID:74904
Oxford Nanopore Ligation Sequencing Kit V14	Oxford Nanopore Technologies	SQK-LSK114
SMRTbell prep kit 3.0	PacificBiosciences	PN:102-141-700
Deposited data		
Okellya marina DNA Nanopore PromethION reads	This study	SRA: SRR35549442
Okellya marina DNA NovaSeq 6000 PE reads	This study	SRA: SRR35549441
Okellya marina RNA NovaSeq 6000 PE reads	This study	SRA: SRR35549440
Okellya sp. SL DNA NovaSeq 6000 PE reads	This study	SRA: SRR35549445
Aegagropila linnaei DNA PacBio Sequel IIe HiFi reads	This study	SRA: SRR35549447
Aegagropila linnaei DNA NovaSeq 6000 PE reads	This study	SRA: SRR35549446
Aegagropila linnaei RNA NovaSeq 6000 PE reads	This study	SRA: SRR35549443, SRR35549444
Chaetomorpha tokyoensis DNA Nanopore PromethION reads	This study	SRA: SRR35549439
Chaetomorpha tokyoensis DNA NovaSeq 6000 PE reads	This study	SRA: SRR35549438
Chaetomorpha tokyoensis RNA NovaSeq 6000 PE reads	This study	SRA: SRR35549437
Okellya marina plastid genome	This study; Mendeley data	https://doi.org/10.17632/684wwdffjh.1
Okellya sp. SL	This study; Mendeley data	https://doi.org/10.17632/684wwdffjh.1
Aegagropila linnaei plastid chromosomes	This study; Mendeley data	https://doi.org/10.17632/684wwdffjh.1
Chaetomorpha tokyoensis plastid chromosomes	This study; Mendeley data	https://doi.org/10.17632/684wwdffjh.1
Experimental models: Organisms/strains		
Okellya marina strain 56A	Australia field sample	Australian National Algal Culture Collection CS-1486
Okellya sp. SL	Sri Lanka field sample	Meise Botanic Garden herbarium specimen BR5010116037406V
Aegagropila linnaei	https://aquafy.com.au/	N/A
Chaetomorpha tokyoensis	Malaysia field sample	N/A
Software and algorithms		
CLC Genomics Workbench	QIAGEN	RRID: SCR_011853; https://digitalinsights.qiagen.com/
Trimmomatic v0.39	Bolger et al.[68]	RRID: SCR_011848; https://github.com/usadellab/Trimmomatic
Flye v.2.8	Kolmogorov et al.[69]	RRID: SCR_017016; https://github.com/mikolmogorov/Flye

(*Continued on next page*)

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SPAdes v3.15.5	Prjibelsk et al.[70]	RRID: SCR_000131; https://github.com/ablab/spades
Canu v2.2	Koren et al.[71]	RRID: SCR_015880; https://github.com/marbl/canu
SLAG v1.0.0	Crane et al.[72]	https://github.com/cfcrane/SLAG
SeqKit v2.5.0	Shen et al.[73]	RRID: SCR_018926; https://bioinf.shenwei.me/seqkit/
Infernal v1.1.5	Nawrocki and Eddy[74]	RRID: SCR_011809; http://eddylab.org/infernal/
CD-HIT v4.8.1	Li and Godzik[75]	RRID: SCR_007105; https://www.bioinformatics.org/cd-hit/
Geneious 2025.0	Kearse et al.[76]	RRID: SCR_010519; https://www.geneious.com/
MetaBCC-LR	Wickramarachchi et al.[77]	https://github.com/anuradhawick/MetaBCC-LR
Minimap2 v2.2	Li[78]	RRID: SCR_018550; https://github.com/lh3/minimap2
Integrated Genomics Viewer v2.17.4	Thorvaldsdóttir et al.[79]	RRID: SCR_011793; https://igv.org/
LRScaf v1.1.10	Qin et al.[80]	https://github.com/shingocat/lrscaf
MFannot	Beck and Lang[81]	https://github.com/BFL-lab/Mfannot
tRNAScan-SE v2.0.9	Chan et al.[82]	RRID: SCR_008637; https://github.com/UCSC-LoweLab/tRNAscan-SE
Mafft v7.520	Katoh and Toh[83]	RRID: SCR_011811; https://mafft.cbrc.jp/
TrimAl v1.4	Capella-Gutiérrez et al.[84]	RRID: SCR_017334; https://trimal.readthedocs.io/
ModelFinder	Kalyaanamoorthy et al.[85]	https://iqtree.github.io/ModelFinder/
IQ-TREE v2.3.3	Minh et al.[86]	RRID: SCR_017254; https://iqtree.github.io/
PhyKIT v1.19.4	Steenwyk et al.[87]	https://github.com/JLSteenwyk/PhyKIT
PartitionFinder2	Lanfear et al.[88]	RRID: SCR_024157; https://www.robertlanfear.com/partitionfinder/
PAL2NAL	Suyama et al.[89]	https://www.bork.embl.de/pal2nal/
PAML	Yang[90]	RRID: SCR_014932; https://github.com/abacus-gene/paml
Bowtie2	Langmead and Salzberg[91]	RRID: SCR_016368; https://github.com/BenLangmead/bowtie2
Samtools	Danecek et al.[92]	https://www.htslib.org/
Freebayes v1.3.6	Garrison and Marth[93]	RRID: SCR_010761; https://github.com/freebayes/freebayes
BEDTools v2.31.1	Quinlan[94]	RRID: SCR_006646; https://bedtools.readthedocs.io/
MeShClust v2	James et al.[95]	https://github.com/BioinformaticsToolsmith/MeShClust
BLAT v34	Kent[96]	RRID: SCR_011919; https://github.com/djhshih/blat
Hisat2	Kim et al.[97]	RRID: SCR_015530; https://daehwankimlab.github.io/hisat2/
MEME suite v5.5.5	Bailey et al.[98]	RRID: SCR_001783
gggenes v0.5.1	Wilkins and Kurtz[99]	https://github.com/wilkox/gggenes
Jalview v2.11.5.1	Waterhouse et al.[100]	RRID: SCR_006459; https://www.jalview.org/
ggseqlogo v0.2	Wagih[101]	https://github.com/omarwagih/ggseqlogo
Tandem Repeats Finder	Benson[102]	RRID: SCR_022065; https://tandem.bu.edu/trf/trf.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Guppy v6.4.6	Oxford Nanopore Technologies	RRID: SCR_023196
ccs v.6.40	PacificBiosciences	RRID: SCR_021174
BLAST+	National Center for Biotechnology Information	https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
R studio 4.4.1	R	https://www.r-project.org/
Jackhmmer	HMMER	http://hmmer.org

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Algae collection and culturing

Collection, culture, and identification of *Okellya marina* strain 56A is described in Wetherbee et al. (2024).[20] *Okellya marina* was collected from an intertidal pool at Narooma Inlet, New

South Wales, Australia (-36.2077222, 150.125111) in March 2019. A clonal culture was established by isolating filaments into K-enriched seawater medium,[65] and maintained in 60 ml at 21 °C under Sylvania 58 Luxline Plus and Gro-Lux fluorescent lamps with a 10:14 hour light:dark cycle. This culture is in the Australian National Algal Culture Collection (ANACC) with accession CS-1486. *Okellya* sp. SL was found as an epiphyte on *Yonagunia ligulata* (Meise Botanic Garden herbarium specimen BR5010116037406V, field ID *Tylotus* sp.) in Sri Lanka (5.95824,80.42267), collected at a depth of < 1m on January 16th, 2003. *Y. ligulata* was preserved in silica gel until DNA extraction and sequencing, from which the presence of *Okellya* sp. SL was discovered. *Aegagropila linnaei* was obtained from the Aquafy online aquarium shop in Australia (https://aquafy.com.au/) in 2023, and cultured in tap water at 27 °C under 30 μmol photons m^{-2} s^{-1} with a 12:12 light-dark cycle. *Chaetomorpha tokyoensis* was collected growing epiphytically on *Avicennia* pneumatophores in Malaysia (4.83568,100.63197) on August 20th, 2014, and cultured in Modified Provasoli's Enriched Seawater[66] at 19-21 °C under 4-10 μmol photons m^{-2} s^{-1} with a 12:12 light-dark cycle. Species confirmation of all algae was done by a maximum-likelihood phylogeny of nuclear 18S genes (Figure S1).

METHOD DETAILS

To aid future research on Cladophorales minicircular genomes, graphical workflows of the more complex or manual methods with additional commentary are provided as Data S1A–S1D.

DNA and RNA extraction and sequencing

DNA extraction for short-read sequencing was performed using the cetyltrimethylammonium bromide (CTAB) protocol optimized for green algae.[67] Briefly, 40mg of freeze-dried algae sample was added to an Eppendorf tube and ground into a fine powder with a micropestle, to which 800 μL of pre-warmed lysis buffer was added (see below for preparation). This was incubated for one hour at 62°C with inversion every few minutes, followed by the addition of 15μL Prot-K (20 mg/mL) and another 50 minutes of incubation and inversion. The temperature was then lowered to 56°C, and 3 μL RNase-A (10 mg/mL) was added and incubated for 30 minutes, followed by a 10 minutes at room temperature. DNA was isolated by adding an equal volume of chloroform:isoamyl alcohol (24:1) followed by centrifugation at 13,000 rpm for 10 minutes, from which the clear upper layer was transferred to a new Eppendorf tube. DNA was precipitated by adding a two-thirds volume of ice-cold 80% isopropanol, inverting and incubating at -20°C for 30 min, which was then centrifuged for 10 minutes at room temperature, and supernatant was discarded. The DNA pellet was washed by adding 1.5 mL 80% ethanol, inverting, and centrifuging to discard supernatant, which was repeated twice, after which the DNA air-dries for 10 minutes. DNA was then dissolved in 50 μL nuclease-free water and incubated for 15 minutes at room temperature.

Lysis buffer:

- 2% CTAB
- 1% polyvinylpyrrolidone
- 0.25% SDS
- 0.25% sodium sulfite (Na_2SO_3)
- 100 mM Tris-HCl
- 1.4 M NaCl
- 20 mM EDTA

RNA was extracted using the Qiagen RNeasy Plant Mini Kit. Library prep and paired-end Illumina 150 bp sequencing was carried out by Novogene (Cambridge, England) for DNA from *Okellya* sp. SL, and by Azenta Life Sciences (Guangzhou, China) for DNA and

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

RNA from *Okellya marina*, *Aegagropila linnaei* and *Chaetomorpha tokyoensis*, all using the Illumina NovaSeq platform. Reads were trimmed of adaptors and low-quality sequences using Trimmomatic v0.39 (ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:2:True LEADING:3 TRAILING:3 SLIDINGWINDOW:4:30 MINLEN:36).[68]

For DNA extraction for long-read sequencing, an adapted phenol:chloroform protocol was used following Nakamura-Gouvea et al. (2022).[103] Briefly, from each algae, 340 mg of material was harvested from fresh culture and snap-frozen in liquid nitrogen. This material was then ground in pestle and mortar in liquid nitrogen into fine powder, and divided into four aliquots. For each, 524 μl of HB buffer and 376 μl of LB buffer (prepared immediately before use, see below for preparation), and 3μl of RNase A (10 mg/mL) was added and mixed by inversion, then incubated on a heated shaker at 1,000 RPM at 65°C for 30 minutes. Following this, 1 ml of phenol:chloroform:isoamyl alcohol in a 25:24:1 ratio was added and mixed by inversion, and centrifuged at 8,000g for 30 minutes at room temperature. Supernatant was transferred to 2 ml Eppendorf tubes, and 100 μl of 3M sodium acetate and 900 μl of pre-cooled 80% isopropanol was added to incubate at -20°C for 1 hour. This was centrifuged at 8,000g for 30 minutes at 4°C, and supernatant was removed. Pellets were washed with 500 μl of pre-cooled 70% ethanol, centrifuged again for 5 minutes, and washed again with 500 μl of pre-cooled 99% ethanol. Supernatant was removed and dried further in a sterile hood for 20 minutes. Finally, 20 μl of nuclease-free water was added to each tube and left to incubate overnight at 4°C. Prior to library preparation, DNA was checked for concentration using a Qubit fluorometer, for purity using NanoDrop A260/280 and A260/230 values, and for fragment size using an Agilent TapeStation 4200.

HB buffer:

- 144 mM Tris
- 36 mM EDTA
- 1584.4 mM NaCl
- 40.6 mM CTAB
- 60.5 mM Na_2SO_3

LB buffer:

- 32.75 mM PVP-40
- 6.2 mM Sarkosyl
- 18.9 mM borax
- 21.4 mM CTAB

For *Aegagropila* PacBio sequencing, libraries were prepared with the SMRTbell prep kit 3.0 and sequenced on the PacBio Sequel IIe platform. Consensus HiFi reads were extracted from subread bam files using ccs v.6.40 (https://github.com/PacificBiosciences/ccs). For *Okellya marina* and *Chaetomorpha* nanopore sequencing, libraries were prepared with the Oxford nanopore Ligation Sequencing Kit (SQK-LSK114) and sequenced on the PromethION platform with FLO-PRO114M flow cells. Raw FAST5 files were base-called with Guppy v6.4.6 software obtained from Oxford nanopore Technologies.

Plastid genome assembly

Nanopore reads of *O. marina* were assembled using Flye v.2.8,[69] with the—meta flag enabled and two polishing iterations. Assembly of Illumina paired-end reads of *Okellya* sp. SL was done with the CLC Genomics Workbench (https://digitalinsights.qiagen.com/) using SPAdes v3.15.5[70] with default parameters.

For assembly of minicircular genomes, a Canu v2.2[71] assembly of *Aegagropila* HiFi reads was first performed, with options stopOnLowCoverage=0 minInputCoverage=0 and various genome sizes attempted. This failed to assemble an identifiable plastid genome. A search for contigs within the estimated plastome GC range (34%–45% GC) found 195 potential contigs, many composed of tandem repeats, and only 4 had detectable plastid genes.

Instead, assembly was conducted with SLAG v1.0.0,[72] which uses an iterative BLAST-based local assembly by searching a query sequence against reads and assembling only matching reads. As we expected genes to be divergent, we sought to use plastid-encoded genes from the species themselves to use as SLAG queries. To do so, long reads were translated into all reading frames with SeqKit v2.5.0[73] and protein sequences from available Cladophorales species were searched against translations using jackhmmer (E-value < 0.001) (hmmer.org). To target rRNA-containing contigs, Infernal v1.1.5[74] was used to build covariance models (commands cmbuild and cmcalibrate) from rRNA sequences of *Boodlea* and *Okellya*, as well as from alignments of Cyanobacteria rRNA genes and Ulvophyceae plastid rRNA genes, and from Rfam bacterial 5S (RF00001), 23S (RF02541) and 16S (RF00017) alignments.[104] All jackhmmer and Infernal alignment regions were extracted from reads and their corresponding nucleotide sequences were clustered with CD-HIT-EST v4.8.1[75] at 97% identity. These sequences were then searched with BLASTx to UniProtKB reference proteomes database and BLASTn to the NCBI nt database to remove bacterial sequences. SLAG was run with each query separately, using the Canu assembler with options -nanopore-raw for nanopore reads and -pacbio-hifi for HiFi reads, and setting stopOnLowCoverage=0 minInputCoverage=0. The OverlapErrorAdjustment.pm file of Canu was also edited to allow assemblies with fewer than 100 reads. Within SLAG, the evalue, secevalue and carryforwardevalue were set to 10^{-5}, extractionoption set to "increment", and a maximum cycle of 20 was used. Resulting contigs were then inspected in Geneious,[76] and any bacterial and mitochondrial sequences assembled were removed by BLASTx and BLASTn searches.

After these initial SLAG assemblies, in attempt to recover as many contigs as possible, sequences that were shared across multiple contigs were identified in Geneious Repeat Finder v1.0.1,[76] with 30 bp minimum repeat length and 20% maximum mismatches, and these common sequences were extracted and used as queries in SLAG. Finally, all assembled contigs were used themselves as queries for a final round of SLAG. All assembled contigs were then clustered together at 90% nucleotide identity with CD-HIT-EST, which takes the longest contigs from the cluster as representative, referred to as nr90 contigs. To estimate the amount of total plastid-derived reads represented in nr90 contig assemblies in *Aegagropila*, MetaBCC-LR v1.0.1[77] was used to cluster all reads into bins with parameters -e tsne -c 100000 -bs 30 -bc 30 -k 5 -s 10. See also Data S1A.

Delimitation of final *Aegagropila* and *Chaetomorpha* chromosomes and confirming circularity

Initial inspections in Geneious[76] suggested that most nr90 contigs over-assembled, as one end of the contig overlapped with the other, suggesting circularity. These were verified by extracting non-overlapping units, concatenating the units and aligning long reads with Minimap2 v2.2[78] to inspect mappings visually in the Integrative Genomics Viewer (IGV)[79] to confirm read coverage across unit boundaries, as well confirm the absence of reads aligning longer than the length of a single unit to both concatenated units and raw assemblies (Figure S2). Out of our original nr90 contigs, some contigs assembled linearly. We attempted to further scaffold these contigs with long reads using LRScaf v1.1.10,[80] which closed a minority contigs assemblies in a circular manner, while most remained linear.

Guided by both IGV mappings and Geneious repeat identification, final nr90 chromosomes were extracted from nr90 contigs. In some instances (particularly in more error-prone nanopore assemblies), nr90 contigs appeared as a long overlapping assembly of shorter mixed reads from similar chromosomes. From these, one or two final chromosomes may have been extracted, depending on regions of read coverage. Extracted chromosomes were independently verified in topology be read mappings as above. All raw assemblies and extracted chromosome sequences and notes on their likely topology are available in Table S2 and at Mendeley Data (DOI: 10.17632/684wwdffjh.1). See also Data S1A.

Annotation of plastid genomes and selection of representative genes

Annotations of plastid genes were made using MFannot,[81] as well as jackhmmer and BLASTx searches against the UniProtKB Swiss-Prot and NCBI nr databases (E-value < 0.001). Protein annotations were manually curated based on homology mappings and ORF predictions in Geneious.[76] rRNA genes were annotated with Infernal searches with Rfam rRNA alignments[74,104] tRNAs were annotated with tRNAscan-SE v2.0.9 using the -O option.[82] Repeats were annotated with Geneious Repeat Finder. *Boodlea* chloroplast sequences and annotations were obtained from NCBI GenBank (accessions MG257795.1 – MG257828.1).

As multiple variants of each gene exists in *Aegagropila*, *Chaetomorpha* and *Boodlea* plastids, the "best" representative for each gene was determined as the copy with the lowest E-value from a BLASTx search to UniProtKB Swiss-Prot and NCBI nr databases. In most representative genes, *de novo* assembled transcripts were found with the same coding sequence. For some genes in *Aegagropila*, transcripts contained better gene sequences than those found in assembled chromosomes, and corresponding sequences were found in individual, unassembled reads. SLAG[72] assembly using these exact coding sequences or reads as queries failed to produce the same sequence. In these cases, gene sequences from individual reads were used as representatives. See Table S1 for details on individual gene annotations.

18S phylogeny for species confirmation

For species confirmation, 18S genes were found in long reads (*O. marina*, *Aegagropila*, Chaetomorpha) or short-read assemblies (*Okellya* sp. SL) with Infernal v1.1.5[74] (Figure S1). 18S sequences of other Cladophorales members were downloaded from the SILVA database,[105] or obtained from references.[19,106,107] Sequences were aligned with MAFFT v7.520 (–auto)[83] and trimmed with trimAl v1.4 (–gt 0.5).[84] ModelFinder[85] implemented in IQ-TREE v2.3.3[86] was used to select GTR+F+R3 as the best fit evolutionary model according to Bayesian Information Criterion. Bootstrap support was estimated using 250 non-parametric bootstrap replicates.

Multi-gene phylogeny

Phylogenetic analysis (Figure 3A) was done using 17 chloroplast protein-coding genes (*atpA*, *atpB*, *atpH*, *atpI*, *petA*, *petB*, *petD*, *psaA*, *psaB*, *psaC*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF* and *rbcL*). Coding sequences of these genes for other green algae and cyanobacteria were obtained from Del Cortona et al. (2017).[14] Genes from *Okellya* and representative genes from *Aegagropila* and *Chaetomorpha* were added to this dataset. Amino acid sequences of each gene were aligned with MAFFT v7.520 (–linsi),[83] trimmed with trimAl v1.4 (-gt 0.8),[84] from which a concatenation and partition file was made with PhyKIT v1.19.4 create_concat.[87] A maximum-likelihood tree was estimated with IQ-TREE v2.3.3,[86] using the generated partition file (-p)[88] and ModelFinder (-m MFP+MERGE),[85] and conducting 200 non-parametric bootstrap replicates.

Estimates of dN/dS in algal genes

Estimated dN/dS rates (Figure 3B) were made on the same dataset used to build the 17-gene ML phylogeny, with some very short sequences from other Cladophorales species removed. Nucleotide sequences were threaded onto amino acid sequences to create codon-aware alignments with PAL2NAL.[89] Internal UGA stop codons were converted to "NNN". The 17-gene ML phylogeny was used with Cladophorales branches labelled as foreground and other Chlorophyta branches labelled as background, with bootstrap

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

values deleted. Rates of non-synonymous and synonymous mutations in each group were estimated with the codeml program from the PAML package.[90] Codeml non-zero control file options were: model = 2, method = 1, codonFreq = 2, kappa = 2, clean data = 1, fix_alpha = 1, omega = 1, ncatG = 10, Small_diff = 1e-8, fix_blength = 1.

Polymorphism quantification on individual chromosomes

For polymorphism quantification on individual chromosomes (Figure 4B), Illumina DNA reads were aligned with Bowtie2 v2.4.4[91] and duplicates were marked and removed with SAMtools v1.21.[92] Coverage of chromosomes was capped at 1000 with jvarkit biostar154220 (https://github.com/lindenb/jvarkit) and variants were called with Freebayes v1.3.6,[93] which to our knowledge is the only variant caller that can handle multiple ploidy levels. Ploidy was set arbitrarily to 10 and minimum observation number set to 3. Variants were filtered by quality $\geq$ 20. SAMtools was used to estimate coverage, and the number of variants called in coding and non-coding regions on chromosomes was normalized by coverage and length of those regions. The first and last 100 bp of chromosomes (as linear representations) were not considered in analysis.

Illumina coverage estimation of *Aegagropila* and *Chaetomorpha* chromosomes and iterative sequence clustering

To estimate coverage of nr90 chromosomes (Figure 5A), Bowtie2 v2.4.4[91] was used to align paired-end Illumina reads to chromosomes, and coverage of each base was determined from bam files with BEDTools v2.31.1 genomeCoverageBed (-bga).[94] Coverage of the entire chromosome was calculated by summing the coverage of all positions, divided by length of the chromosome. Then, coverage of chromosomes in both *Aegagropila* and *Chaetomorpha* was normalized relative to each other by dividing by total number of reads aligned to all chromosomes in each species (24,997,163 in *Aegagropila* and 15,892,082 in *Chaetomorpha*) and multiplying by 10^6.

Iterative clustering by decreasing sequence identity thresholds (Figure 5A) was accomplished with MeShClust v2,[95] which can accommodate nucleotide clustering at low identities. After a clustering round, the "center" chromosome of each cluster (assigned by MeShClust v2) were taken, which the next round of clustering would be performed on. Cluster assignment of each member of this second round would then be assigned to all other members of the cluster it belonged to in the previous round. This was done because clustering cannot be performed group-wise, and re-clustering of all chromosomes at each step results in differences in cluster assignments at different thresholds. See also Data S1A.

Transcriptome assembly, identification of gene variant expression and RNA editing

Illumina paired-end RNA reads of *Okellya marina*, *Aegagropila* and *Chaetomorpha* were assembled with rnaSPAdes v3.15.5[70] using options -k 21,33,55,77,99,121. Assembled transcripts were aligned to nr90 chromosomes using BLAT v34[96] with options -t=dna -q=rna -oneoff=2 -minScore=10. Transcripts that had at least 90% of their length aligned to chromosomes were extracted and encoding genes were identified by jackhmmer (hmmer.org) search of six-frame translations to the UniProtKB Swiss-Prot database (E-value < 0.001). Transcripts containing plastid-encoding genes were extracted and re-aligned to nr90 chromosomes in Geneious[76] using Minimap2 v2.2[78] long-read spliced alignment. If more than one transcript encoding the same gene aligned to chromosomes, these were inspected for differences in coding sequences in Geneious. If a long, gene-containing transcript was not identical to chromosomes, its sequence was searched for in reads by BLASTn to confirm it was not a result of editing.

Additionally, RNA and DNA Illumina reads were aligned to all assembled chromosomes of *Aegagropila* and *Chaetomorpha*, clustered at 99% identity (instead of nr90 chromosomes). DNA reads were aligned with Bowtie2 v2.4.4[91] and RNA reads with Hisat2 (–sp 10,3—pen-noncansplice 6—max-intronlen 1000).[97] Duplicates were marked and removed with SAMtools.[92] Variants were called with Freebayes v1.3.6,[93] with ploidy set to 10 and minimum observation number set to 3. Variants were filtered by quality $\geq$ 20. See also Data S1B.

Motif identification across minicircular chromosomes

Common motifs on nr90 chromosomes (Figure 5B) were found in each genome with MEME v5.5.5 (-minsites 10 -maxsites 300 -maxw 300 -nmotifs 20) from the MEME suite.[98] Motifs were searched back against the genomes with FIMO of the MEME suite and matches with q-value < 10^{-10} were extracted. Pairwise identity of each motif across all instances on all chromosomes was calculated with PhyKIT v1.19.4 pairwise_identity.[85] Smaller motifs were found recurring together in groups, and for better visualizing these assigned groups were plotted in Figure 5B. Motif annotations were plotted with R package gggenes v0.5.1.[99]

Estimation of internal UGA reassignments

To estimate the amino acid internal UGA codons code for, amino acid alignments used in the 17-gene ML phylogeny were subsetted to selected green algae (Figure 6A). These alignments were then subsetted to only positions where a UGA codon was present in a Cladophorales member, and visualized with Jalview v1.8.[100] Sequence logos of these positions (Figure 6B) across other green algae were made with the R package ggseqlogo v0.2,[101] setting method = "probability".

As Cladophorales genes are very divergent, to consider alignment quality around UGA positions and the possibility of using non-functional genes in alignments, we collected all possibly functional copies of each gene (no internal UAA/UAG stop codons, long open reading frames, no frame-shift mutations). Amino acid alignments were remade for all genes without Cladophorales. For each species, each potential protein was added to the relevant Chlorophyta alignment with MAFFT v7.520 linsi—add.[83] The average percent identity of alignment of the added protein compared to all other proteins was calculated with PhyKIT v1.19.4

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

pairwise_identity.[87] Likewise, for each UGA position on the added protein the average percent identity of a 10 AA window around that position was calculated. For each UGA, the frequency of all other amino acids in that position across Chlorophyta was calculated. These metrics were plotted in Figure 6C. See also Data S1D.

Search for trans-splicing of fragmented rRNA genes

As rRNA genes are present as fragments on separate chromosomes, a search for potential trans-splicing was performed by inspecting *de novo* assembled RNA transcripts for longer rRNA sequences. Ribosomal RNA gene fragments from assembled chromosomes were extracted and used as BLASTn queries to transcripts. On transcript hits, rRNA gene regions were better annotated by searches with Infernal v1.1.5,[74] as was done to annotate original rRNA fragments on chromosomes. These were then used as BLASTn queries against the SILVA database[101] to remove bacterial sequences. Non-bacterial rRNA sequences were examined for rRNA Infernal predictions longer in length than those found in chromosomes, but none were found.

Search for homoconcatemers and heteroconcatemers in reads

A search for homoconcatemers of chromosomes in reads was conducted, as homoconcatemers may be evidence of rolling-circle replication or recombination, while heteroconcatemers would indicate recombination between minicircles. In *Okellya marina*, nanopore reads were mapped to the plastid with Minimap2 v2.21.[78] Only 11 reads mapped longer than the plastid, which were visually inspected in Geneious.[76] For *Aegagropila* and *Chaetomorpha*, all reads > 5 kb with GC within the plastid ranges (35%–46% for Aegagropila, 49%–55% for Chaetomorpha) were inspected. Reads were masked with Tandem Repeats Finder (options 2 5 7 80 10 50 2000 -l 10 -h -d -m),[102] and reads with < 40% tandem repeats were used, unmasked.

To search for homoconcatemers, reads were mapped back against each other with BLASTn. Reads that mapped back to themselves with length > 2 kb in two non-overlapping locations with the same orientation (indicating head-to-head mappings) were retrieved. These were then mapped back to all assembled chromosomes with BLASTn, and reads that aligned >2 kb to a plastid chromosome were selected. These were visually inspected in Geneious for identical repeating regions that would indicate homoconcatemers (i.e. two or more long, directly repeating regions covering the read), of which none were found.

To search for heteroconcatemers, reads were mapped to all assembled chromosomes with BLASTn. Reads that aligned with length >2 kb to two separate chromosomes were retrieved. These were then used to search against all assembled chromosomes in Geneious with Minimap2, and alignments were inspected visually for reads that mapped different assembled chromosomes to different portions of the read, of which none were found. See also Data S1C.

Identification of plastid-mapping reads with inverted structure

To search for reads of plastid origin with large inverted repeats, long reads were mapped to all assembled chromosomes of *Aegagropila* and *Chaetomorpha* using Minimap2 v2.21.[78] Mapped reads were extracted, and masked with Tandem Repeat Finder (options: 2 5 7 80 10 50 2000 -l 10 -h -d -m).[102] Reads with < 40% tandem repeats were used, and mapped back to themselves with BLASTn. Reads that aligned with > 300 bp to themselves in opposite orientations were extracted, and again searched against plastid nr90 chromosomes with BLASTn. Reads that aligned > 1000 bp to plastid chromosomes were then extracted. These were inspected visually in Geneious[76] for long inverted repeats with Geneious Repeat Finder (maximum 20% divergence), and for plastid mappings to inverted areas by Minimap2,[78] of which 51 in *Aegagropila* and 96 in *Chaetomorpha* fit this criteria. Example reads and annotations were plotted with R package gggenes v0.5.1[99] (Figure 5C). See also Data S1C.

Identification of reads with plastid sequences and retrotransposon-type elements

During our various mappings of reads back to plastid chromosomes, we became aware of long reads mapped that contained retrotransposon-like gene fragments, as was found by the authors of the *Boodlea* plastome.[14] To search specifically for these reads, long reads were mapped to *Okellya marina*, *Aegagropila* and *Chaetomorpha* plastid chromosomes with Minimap2.[78] For *Aegagropila* and *Chaetomorpha*, reads that aligned to at least 25% of the chromosome length but were longer than the chromosome itself were extracted, while in *O. marina* reads > 10 kb in length with > 2000 bp mapped to the plastid were extracted. These were then translated to all six reading frames with SeqKit v2.5.0,[73] and searched against the UniprotKB Swiss-Prot database with jackhmmer (hmmer.org). These were then searched for hits to retrotransposon polyprotein sequences with a minimum E-value of 10^{-5}. A total of 34 reads in *O. marina*, 93 reads in *Aegagropila* and 13 in *Chaetomorpha* were found this way. Selected reads with annotations were plotted with R package gggenes v0.5.1[99] (Figure S6).

QUANTIFICATION AND STATISTICAL ANALYSIS

For estimation of plastid-derived reads represented in nr90 contig assemblies in Aegagropila, MetaBCC-LR v1.0.[77] was used to cluster all reads into bins with parameters -e tsne -c 100000 -bs 30 -bc 30 -k 5 -s 10.

For the 18S maximum-likelihood phylogeny, ModelFinder[85] implemented in IQ-TREE v2.3.3[86] was used to select GTR+F+R3 as the best fit evolutionary model according to Bayesian Information Criterion. Bootstrap support was estimated using 250 non-parametric bootstrap replicates. For the phylogeny built from plastid genes, a partition file (-p)[88] was generated and ModelFinder was used to find to the best-fit models for each partition under the Bayesian Information Criterion (-m MFP+MERGE).[85] Bootstrap support was estimated using 200 non-parametric bootstrap replicates. For estimation of dN/dS rates, CODEML[90] non-zero control file

Please cite this article in press as: Bjornson et al., Stepwise evolutionary transitions in Cladophorales plastid genomes reveal origins of hairpin chromosomes, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.02.038

options were: model = 2, method = 1, codonFreq = 2, kappa = 2, clean data = 1, fix_alpha = 1, omega = 1, ncatG = 10, Small_diff = 1e-8, fix_blength = 1.

For polymorphism quantification on chromosomes, Illumina read coverage was capped at 1000 and the number of variants were normalized by coverage and length of regions under inspection. To estimate Illumina read coverage on all chromosomes, coverage of the entire chromosome was calculated by summing the coverage of all positions, divided by length of the chromosome, divided by the total number of reads aligned to all chromosomes in each species and multiplied by 10^6.

Additional statistical details of analyses including E-values and software parameters used are given in figure legends or in method details.