Molecular Analysis of Parasites in the Choreocolacaceae (Rhodophyta) Reveals a Reduced *Harveyella mirabilis* Plastid Genome and Supports the Transfer of Genera to the Rhodomelaceae (Rhodophyta)

Eric D. Salomaki[2]

Department of Biological Sciences, University of Rhode Island, Kingston, RI, USA.

Gary W. Saunders

Centre for Environmental and Molecular Algal Research, Department of Biology, University of New Brunswick, Fredericton, NB, E3B 5A3, Canada.

Christopher E. Lane

Department of Biological Sciences, University of Rhode Island, Kingston, RI, USA.

[1]Received.

[2]Author for correspondence: email eric.salomaki@gmail.com.

Running Title: *Alloparasite plastid from Harveyella mirabilis*

Abstract:

Parasitism is a life strategy that has repeatedly evolved within the Florideophyceae. Until recently, the accepted paradigm of red algal parasite evolution was that parasites arise by first infecting a close relative and, either through host jumping or diversification, adapt to infect more distant relatives. The terms adelphoparasite and alloparasite have been used to distinguish parasites that are closely related to their hosts from those more distantly related to their hosts, respectively. Phylogenetic studies have cast doubt on the utility of these terms as data show that even alloparasites predominately infect with the same family. All adelphoparasites that have been investigated have lost a native plastid and instead hijack and incorporate a copy of their hosts' plastid when packaging spores. In contrast, a highly reduced plastid that has lost all genes involved with photosynthesis was sequenced from the alloparasite *Choreocolax polysiphoniae*, which indicates that it did not pass through an adelphoparasite stage. In this study we investigate whether other species in the Choreocolacaceae, a family of alloparasites, also retains its native plastid, as well as test the hypothesis that alloparasites can arise and subsequently speciate to form monophyletic clades that infect a range of hosts. We present the plastid genome for *Harveyella mirabilis* which, similar to that of *C. polysiphoniae*, has lost genes involved in photosynthesis. The *H. mirabilis* plastid shares more synteny with free-living red algal plastids than that of *C. polysiphoniae* Phylogenetic analysis identifies a well-supported monophyletic clade of parasites in the Choreocolacaceae, which retain their own plastid genomes, within the Rhodomelaceae. We therefore transfer genera in the Choreocolacaceae to the Rhodomelaceae.

Keywords: Alloparasite, Adelphoparasite, *Choreocolax*, *Leachiella*

INTRODUCTION:

Since the 1845, when *Microcolax botryocarpa* Hooker & Harvey, became the first formally described parasitic red alga, the known biodiversity of red algal parasites has been steadily increasing (Reinsch 1875, Setchell 1914, 1923, Goff 1982, Preuss et al. 2017). Red algal parasites exclusively infect other rhodophytes and are predominately unpigmented, appearing as galls or irregular growths on their free-living red algal hosts. Recent counts have identified 121 distinct species of red algal parasites distributed across eight different orders of the Florideophyceae (Salomaki and Lane 2014, Blouin and Lane 2015, Preuss et al. 2017). Florideophyceae seem to be prone to adopting parasitism, in large part due to their ability to form direct cell-to-cell fusions between adjacent, non daughter cells (Wetherbee and Quirk 1982, Goff and Coleman 1985). With few exceptions [e.g. *Choreonema thuretii* (Broadwater and LaPointe 1997)], red algal parasites leverage their ability to form cell fusions as a means of infecting their host (Goff and Coleman 1985, Zuccarello et al. 2004).

Two main hypotheses have been proposed for the origin of red algal parasites. Setchell (1918) initially proposed that parasites arose as spores from their host, which had mutated to no longer be capable of photosynthesis. Sturch (1926) later proposed that parasites evolved from epiphytes that penetrated and, over time, became nutritionally reliant on the host. Feldmann and Feldmann (1958) later added to Sturch's hypothesis by suggesting that epiphytes that are closely related to their hosts are more likely to succeed in forming secondary pit connections and therefore increase the likelihood of successfully establishing a parasitic relationship.

66 Despite their diminutive habit, parasitic red algae share morphological
67 characteristics with other close relatives, and thus were assigned to tribes or families at
68 the time of their initial discovery (Reinsch 1875, Feldmann and Feldmann 1958).
69 Historically parasites that have infected close relatives have been considered
70 'adelphoparasites', while those more distantly related to their hosts are called
71 'alloparasites' (Feldmann and Feldmann 1958, Goff et al. 1996). Approximately 80% of
72 the described red algal parasite diversity is considered to be adelphoparasitic, while the
73 remaining 20% is alloparasitic (Goff 1982, Salomaki and Lane 2014, Blouin and Lane
74 2015). Based initially on morphological observations, and later coupled with molecular
75 data, it was proposed that parasites evolve sympatrically with their hosts, as
76 adelphoparasites, and over time diversify or adapt to infect new and more distantly
77 related hosts, becoming alloparasites (Feldmann and Feldmann 1958, Goff et al. 1997).
78 Sturch (1926) initially described the family Choreocolacaceae as a family in the
79 Gigartinales, which consisted of morphologically reduced parasites lacking chlorophyll
80 including members of the genera *Choreocolax, Harveyella,* and *Holmsella*. The
81 Choreocolacaceae was the subject of a thorough phylogenetic analysis of alloparasites to
82 confirm whether the family was truly a monophyletic clade of parasitic red algae
83 (Zuccarello et al. 2004). This study supported previous morphological observations
84 (Fredericq and Hommersand 1990) confirming that the genus *Holmsella* was a member
85 of the Gracilariaceae and questioned the legitimacy of recognizing a family of red algal
86 parasites.
87 In the few other cases where molecular tools have been applied to assess
88 evolutionary histories of red algal parasites, data suggests that red algal parasites arise

89 though independent evolutionary events (Goff et al. 1996, 1997, Kurihara et al. 2010). In
90 addition to phylogenetic analyses, molecular tools have also been applied to investigate
91 the parasite-host dynamics throughout parasite development. Analyses of the
92 adelphoparasites *Gardneriella tuberifera* Kylin*, Gracilariophila oryzoides* Setchell & H.
93 L. Wilson demonstrated that, although the parasites maintain a native mitochondrion,
94 they have lost their native plastid and instead 'hijack' a host plastid when packaging their
95 own spores (Goff and Coleman 1995). To date, all red algal parasites examined maintain
96 a fully functional mitochondrion (Salomaki and Lane 2016). All adelphoparasites that
97 have been investigated have lost their native plastid (Goff and Coleman 1995; Salomaki
98 and Lane 2014; Salomaki and Lane unpublished). In contrast, a highly reduced native
99 plastid was sequenced from the alloparasite *Choreocolax polysiphoniae* Reinsch, which
100 has lost genes involved in photosynthesis, yet maintains functions including fatty acid
101 and amino acid biosynthesis (Salomaki et al. 2015). The lack of plastids in
102 adelphoparasites, in combination with finding a native plastid in the alloparasite *C.*
103 *polysiphoniae*, demonstrates that not all parasites pass through an adelphoparasite stage
104 and that there are multiple paths to parasitism in red algae.
105 In their study examining relationships in the Choreocolacaceae, Zuccarello *et al.*
106 (2004) found that *Holmsella pachyderma* and *Holmsella australis* form a monophyletic
107 clade within the Gracilariaceae. Additionally they identify that parasite genera
108 *Choreocolax*, *Harveyella*, and *Leachiella* are members of the Rhodomelaceae, though
109 have fairly low support for their relationships to each other. Using molecular data we
110 investigate the relationships of *Choreocolax*, *Harveyella*, and *Leachiella*, and test the
111 hypothesis that alloparasites can arise and subsequently speciate, forming monophyletic

112 clades that infect a range of hosts. Furthermore, we set out to determine if another
113 alloparasite besides *C. polysiphoniae* retains its own native plastid.
114
115 **MATERIALS AND METHODS**
116 *Sample Collection and DNA Extraction*
117 *Choreocolax polysiphoniae*, found on its host, *Vertebrata lanosa* (Linneaus) T.A.
118 Christensen, was collected at Beavertail State Park, Jamestown, RI, USA. *Harveyella*
119 *mirabilis* (Reinsch) F.Schmitz & Reinke, on its host, *Odonthalia washingtoniensis* Kylin,
120 was collected from Cattle Point, Friday Harbor, WA, USA (48.452428, -122.962774),
121 and *Leachiella pacifica* Kugrens, on *Polysiphonia hendryi* N.L. Gardner, was collected
122 off the dock at Friday Harbor Laboratories, Friday Harbor WA, USA (48.545066, -
123 123.012268). Individual parasite galls from *Harveyella mirabilis* and *Leachiella pacifica*
124 were excised from their respective hosts and collected in a 1.5 mL microcentrifuge tube.
125 The parasite tissue was hand-ground using a Corning Axygen® PES-15-B-SI disposable
126 tissue grinder pestle in a 1.5 mL microcentrifuge tube while submerged in 100µL of
127 DNA extraction buffer (Saunders 1993). DNA was extracted from specimens using a
128 standard phenol/chloroform extraction with all ratios adjusted for an initial buffer volume
129 of 100µL (Saunders 1993). Additionally, lyophilized DNA from specimen GWS021225
130 (Saunders 2014) was acquired from the Saunders lab and rehydrated in 20µL of 5mM
131 Tris/HCl pH 8.5.
132
133 *Molecular analyses*

A 664 bp fragment at the 5-prime end of the mitochondrial cytochrome oxidase 1 gene (COI-5P), which has been used extensively for barcoding red algal species (Saunders 2005), was PCR amplified using the GWSFn (Le Gall and Saunders 2010) and GWSRx (Saunders and Moore 2013) primer pair according to protocols in Saunders and Moore (2013). All PCR reactions were completed with Ex Taq ^{TM} DNA Polymerase (Takara Bio USA, Mountainview, CA, USA) following the manufacturer's recommendations. The BigDye Terminator v3.1 Cycle Sequencing Kit (PE Applied Biosystems [ABI], Foster City, CA, USA) was used for the sequencing of PCR products, and reactions were analyzed using an ABI Prism 3130XL genetic analyzer. Sequences were trimmed and assembled using Geneious v. 9.1.5. The consensus sequences resulting from the assembly were BLAST searched against the NCBI GenBank database and the Barcode Of Life Database (boldsystems.org) to find similar publicly available sequences.

In addition to PCR amplification of the COI-5P, genomic DNA was amplified from all specimens using the illustra Single Cell GenomiPhi DNA Amplification kit (GE Healthcare Life Sciences, Pittsburgh, Pa) according to manufacturer protocols. Whole-genome amplification of an individual *C. polysiphoniae* parasite was completed using the illustra GenomiPhi Hy DNA Amplification Kit according to manufacturer protocols. The COI-5P was PCR amplified and sequenced, as previously described, from the whole-genome amplified DNA to screen for contamination resulting from the amplification process. Libraries for Illumina sequencing were constructed from the amplified DNA on the Apollo 324 robot using the PrepX ILM DNA Library Kit (Wafergen Biosystems, Freemont, CA, USA). The *Harveyella mirabilis, Leachiella pacifica,* and GWS021225 libraries were multiplexed and sequenced on an Illumina MiSeq paired-end 250 x 250

157 basepair run. The *C. polysiphoniae* library was sequenced on separate full-cell Illumina
158 MiSeq paired-end 300 x 300 basepair run. The sequencing effort resulted in 29,355,470
159 paired-end reads for *C. polysiphoniae,* 7,923,094 paired end reads for *Harveyella*
160 *mirabilis*, 7,502,360 paired-end reads for *Leachiella pacifica*, and 8,869,740 paired end
161 reads for GWS021225. For all datasets, sequences with PHRED scores <30 were
162 removed and the remaining reads were trimmed of adapter sequences. Additionally,
163 fifteen 5' and five 3' nucleotides were trimmed from the remaining reads and all reads
164 under 100 nucleotides were removed from the dataset. All trimming was completed using
165 CLC Genomics Workbench v. 9.5.2 (CLC Bio-Qiagen, Aarhus, Denmark) and the
166 remaining reads were assembled using default parameters in CLC Genomics Workbench
167 v. 9.5.2.
168
169 *Phylogenetic Analysis*
170 A total of 823 Rhodomelaceae COI-5P sequences available on GenBank were
171 downloaded into Geneious v. 9.1.5 {FormattingCitation} and aligned with the data
172 generate by PCR for *Harveyella mirabilis* and *Leachiella pacifica.* This alignment was
173 subjected to maximum likelihood analysis using RAxML with the GTR+I+G model with
174 1,000 bootstrap replicates. Based upon the results of the preliminary tree, 18S, 28S, COI-
175 5P, and *rbc*L data were compiled from GenBank for thirteen of the closest relatives to a
176 weakly supported clade containing *Choreocolax polysiphoniae* Reinsch, *Harveyella*
177 *mirabilis, Leachiella pacifica*, and GWS021225 (Table 1). The longest sequence for each
178 gene from this compiled dataset was used as a query against local BLAST databases
179 created from the *C. polysiphoniae, H. mirabilis, L. pacifica,* and GWS021225 assemblies,

(Table 1). The 18S and 28S regions were extracted from the ribosomal cistrons identified in each assembly based upon its alignment to the query gene.

Each dataset was aligned using he MAFFT plugin in Geneious v. 9.1.5 and subsequently concatenated using Sequence Matrix v. 1.7.8 {FormattingCitation}. The concatenated dataset was then subjected to phylogenetic analysis using maximum likelihood, implemented in RAxML v. 8.2.2 {FormattingCitation} and Bayesian Inference using MrBayes v. 3.2.2 {FormattingCitation}. For the maximum likelihood analysis the GTR+I+G model was used and bootstrap support values were calculated using 1,000 bootstrap replicates. For the Bayesian inference analysis, two Metropolis-coupled Markov chain Monte Carlo (MCMCMC) runs consisting of one cold chain and three hot chains were preformed. Each run was sampled every 100 generations for 1,010,000 generations. After confirming the runs converged by checking to ensure that the average standard deviation of split frequencies was below 0.01, the trees were merged. The resulting tree and posterior probabilities were calculated from the 20,202 trees generated.

Plastid Genome Annotation

A 90,654 bp contig was identified as the plastid genome of *Harveyella mirabilis* from the previously described assembled Illumina MiSeq data. Open reading frame (ORF) prediction on the *H. mirabilis* plastid was done in Geneious Pro v6.1 and the resulting ORFs were manually annotated using GenBank and Pfam (Finn et al. 2010, 2015) databases. Functional annotations were assigned from the UniProt (The UniProt Consortium 2017) and KEGG databases (Kanehisa et al. 2016). Genes found in red algal

plastid genomes that were missing from the *H. mirabilis* plastid were searched for using BLAST, against the plastid sequence and the genomic assemblies to verify their absence and check for evidence of transfer to another genetic compartment. The plastid genome sequence was submitted to the tRNAscan-SE online server v1.21 (Schattner et al. 2005) for identification of tRNA sequences and to MFannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl) to identify rRNA sequences and confirm manual gene annotations.

Plastid Genome Comparative Analysis

The circular plastid genomes of the free-living florideophytes, *Calliarthron tuberculosum* (KC153978), *Ceramium cimbricum* (KR025491), *Ceramium japonicum* (KX284719), *Chondrus crispus* (HF562234), *Dasya binghamiae* (KX247284), *Laurencia sp*. JFC0032 (LN833431), *Grateloupia taiwanensis* (KC894740), *Gracilaria tenuistipitata* (AY673996), *Vertebrata lanosa* (KP308097), and the alloparasites *Choreocolax polysiphoniae* (KP308096) and *Harveyella mirabilis* (XXXXXXXX) were arranged so their sequences began with the *fts*H gene. Whole genome alignment was completed using the default settings for the progessiveMauve algorithm in the Mauve v2.3.1 (Darling et al. 2004).

RESULTS

Identification of a Cryptic Parasite

A maximum likelihood phylogenetic analysis of 823 Rhodomelaceae COI-5P sequences was completed using RAxML. This analysis provided weak support (bootstrap support of 23) for a clade containing the alloparasites *Choreocolax polysiphoniae, Harveyella mirabilis, Leachiella pacifica*, as well as 2 sequences labeled *Rhodomela* sp1Cal voucher GWS021225 (KM254767) and *Rhodomela* sp1Cal voucher GWS021347 (KM254267) (Supplemental Figure 1). Interestingly, these two GenBank sequences had bootstrap support of 100 as being sister to the sequence of *Leachiella pacifica* generated in this study (data not shown). Based upon these results, the specimens were reexamined and parasite galls were found on the host, which was subsequently identified as *Polysiphonia paniculata* based upon *rbc*L sequence data.

Monophyletic Alloparasite Clade

Although identification of parasites on the samples GWS021225 and GWS021347 was a direct result of resolving a clade of parasite sequences, bootstrap support for a monophyletic clade of alloparasites was weak and a subsequent Bayesian phylogeny failed to recover the same clade (data not shown). Based upon this original analysis, COI-5P data from a reduced dataset of the parasites and 28 of their closest relatives was subjected to maximum likelihood analysis, which again recovered a poorly supported monophyletic alloparasite clade (Supplemental Figure 2). To resolve the issue of low statistical support, thirteen taxa that were continually allied to the alloparasite clade were selected based upon availability of published comparative data. Phylogenetic analysis was conducted on a concatenated dataset comprised of 18S, 28S, COI-5P and *rbc*L (Table 1) from the 13 free-living species, and the alloparasites *C. polysiphoniae, H.*

mirabilis, L. pacifica, and the parasite on GWS021225 (from here on called *Leachiella* sp.). With the additional data beyond the COI-5P, both maximum likelihood and Bayesian inference recovered a monophyletic clade of red algal alloparasites with 75% bootstrap support and 0.99 posterior probability (Figure 1). *Choreocolax polysiphoniae* was the earliest branching parasite followed by *Harveyella mirabilis*, with *Leachiella pacifica* and *Leachiella* sp. being recovered as sister taxa with 100% bootstrap support/1.0 posterior probability (Figure 1).

To exclude long-branch attraction as a contributor to the monophyly of the clade of alloparasites, maximum likelihood analyses were conducted individually on the 18S and 28S gene trees, as well as the concatenated dataset with both species of *Leachiella* (the longest branches) removed. All analyses recovered a monophyletic clade of alloparasites, however the branching patterns differed between them. In the 18S gene tree, the clade had 82% bootstrap support, however *H. mirabilis* was the earliest branching parasite followed by *L. pacifica, L.* sp., and *C. polysiphoniae*. The 28S gene tree recovered a monophyletic parasite clade with the same branching pattern observed in Figure 1, but with only 65% bootstrap support. The concatenated analysis with the two species of *Leachiella* removed also recovered *C. polysiphoniae* and *H. mirabilis* as sister species with 70% bootstrap support.

Taxonomic Considerations

Choreocolax polysiphoniae was initially described by Reinsch (1875) from the Atlantic coast of North America as parasitic on *Polysiphonia fastigiata* (now *Vertebrata lanosa*). Specimens used the phylogenetic analysis presented here (Figure 1) were

271 collected at Beavertail State Park in Jamestown, RI, USA infecting *Vertebrata lanosa* and
272 are a strong match to the type description (Reinsch 1875). As the type collection cannot
273 be located, we formally lectotypify *C. polysiphoniae* on image #49 accompanying the
274 description in Reinsch (1875). Based upon the molecular analyses here, which resolved a
275 monophyletic clade of parasitic red algae within the Rhodomelaceae (Figure 1), we
276 formally transfer *Choreocolax* to the Rhodomelaceae and recognize Choreocolacaceae as
277 a synonym of this family. To adhere to the principle of monophyly, the genus *Harveyella*,
278 based on the type and only species *H. mirabilis* and included in our analyses (Figure 1), is
279 also transferred to the Rhodomelaceae.
280
281 *Harveyella mirabilis Plastid Genome*
282 The plastid genome of *Harveyella mirabilis* was assembled as a 90,654 kb
283 circular molecule with 322x coverage. The plastid genome has an overall AT content of
284 76.5% and contains 84 protein coding genes, 3 rRNAs, and 23 tRNAs (Figure 2). Similar
285 to the *Choreocolax polysiphoniae* plastid (Salomaki et al. 2015), all genes related to
286 photosynthesis have been lost with the exception of *pet*F which has been demonstrated to
287 be involved in electron transport in other metabolic pathways (Happe and Naber 1993,
288 Jacobs et al. 2009). Genes involved in transcription/translation and fatty acid, amino acid,
289 protein, isoprene biosynthesis remain conserved. As in the *C. polysiphoniae* plastid, *glt*B
290 appears to be a vanishing pseudogene. BLAST similarity searches are able to find
291 conserved homology, however the presence of stop-codons throughout the region
292 suggests that the gene is likely no longer capable of being completely translated.
293

294 *Plastid Genome Comparisons*

295 A whole genome MAUVE alignment of the *H. mirabilis* with *C. polysiphoniae* and nine
296 representative free-living Florideophyceae plastid genomes identified 13 locally collinear
297 blocks in the *H. mirabilis* genome that aligned with the free-living plastids (Figure 3).
298 There were no rearrangements or inversions in the *H. mirabilis* plastid genome when
299 compared to photosynthetic Rhodomelaceae taxa. When aligning *H. mirabilis* to the *C.*
300 *polysiphoniae* plastid genome, 11 locally collinear blocks are identified, and several
301 genome inversions and rearrangements are evident (Figure 4). Additionally, gene content
302 varies slightly between the two parasite plastid genomes. *Harveyella mirabilis* retains
303 *arg*B, *car*A, *inf*B, *rnz*, *rpl*9, *rpl*17, *rpl*24, *rpl*32, *rpl*33, *rpl*34, *rpo*Z, *rps*18, *rps*20, *ycf*21,
304 which have all been lost in *C. polysiphoniae* (Table 2), while *C. polysiphoniae* maintains
305 copies of *dna*B and *fab*H which have both been lost in the *H. mirabilis* plastid genome.
306
307 **DISCUSSION**
308 *Origin of Parasites*
309 With 121 described parasites occurring across eight Florideophyeceae orders
310 (Salomaki and Lane 2014, Blouin and Lane 2015, Preuss et al. 2017), red algae appear
311 more able to transition from autotrophy to parasitic lifestyles than any other eukaryotic
312 lineage. The terms adelphoparasite and alloparasite have traditionally been used to
313 describe parasites that infect hosts within their tribe/family, or in different tribes/families,
314 respectively. The use of these terms has been questioned as molecular data have revealed
315 that alloparasites, like adelphoparasites, infect close relatives rather than distantly related
316 species (Zuccarello et al. 2004, Kurihara et al. 2010). Phylogenetic analysis using data

317 presented here place *Choreocolax polysiphoniae*, *Harveyella mirabilis*, and two species
318 of *Leachiella* firmly within the same family as their hosts further supporting
319 abandonment of these terms for differentiating two types of red algal parasites.
320
321 *The Demise of Exclusively Parasitic Families*
322 Sturch (1926) initially described the Choreocolacaceae as a family of
323 holoparasites, containing the genera *Choreocolax, Harveyella,* and *Holmsella*. However,
324 more recent morphological investigation considered that the genus *Holmsella* was related
325 to the parasites *Gelidiocolax* and *Pterocladiophila*, and it was moved to the family
326 Pterocladiophilaceae in the Gracilariales (Fredericq and Hommersand 1990). Their
327 observations were subsequently supported by molecular data generated with the specific
328 aim of testing the phylogenetic affinities of parasites that Sturch had assigned to the
329 Choreocolacaceae. This work demonstrated that *Holmsella australis* and *Holmsella*
330 *pachyderma* formed a well supported monophyletic clade within the Gracilariaceae
331 (Zuccarello et al. 2004). Their molecular data also indicated that *Choreocolax* and
332 *Harveyella* belong in the Ceramiales, leading the authors to question whether the
333 Choreocolacaceae should continue to be recognized (Zuccarello et al. 2004). However,
334 their use of 18S sequence data was insufficient to resolve the issue of a monophyletic
335 clade for these parasites. Other authors have also noted that species in *Choreocolax,*
336 *Harveyella*, and *Leachiella* have features aligning them to the Ceramiales, but again,
337 taxonomic affinities among the parasite species and within this order remained uncertain
338 (Goff and Cole 1975, Kugrens 1982, Fredericq and Hommersand 1990). Our data
339 confirm the findings of Zuccarello, Moon, and Goff (2004), placing *Choreocolax*,

340 *Harveyella* and *Leachiella*, within the Rhodomelaceae. Furthermore, the phylogeny
341 utilizing additional molecular markers provides strong support for a monophyletic clade
342 containing *Choreocolax*, *Harveyella* and *Leachiella* (Figure 1), and supports the placing
343 the Choreocolacaceae in synonymy with the Rhodomelaceae.
344
345 *Cryptic Species*
346 Until recently, the dogma behind red algal parasite evolution was the notion that
347 parasites arise sympatrically as adelphoparasites and over time evolve and adapt to infect
348 more distant hosts becoming alloparasites (Goff et al. 1996, 1997). Based on molecular
349 data demonstrating that alloparasites also infect members of their own family (Zuccarello
350 et al. 2004, Kurihara et al. 2010), we now consider that these terms are not suitable for
351 distinguishing parasites. The recognition of multiple monophyletic parasite clades, the
352 *Holmsella* clade (Zuccarello et al. 2004), and the Rhodomelaceae clade uncovered here
353 containing *Choreocolax, Harveyella,* and *Leachiella* (Figure 1), supports the idea of a
354 sympatric origin of alloparasites with subsequent speciation as parasites adapt to their
355 hosts, without passing through an adelphoparasite-like stage.
356
357 Previously, *Leachiella pacifica* was described from multiple hosts including
358 members of the genus *Polysiphonia* and *Pterosiphonia* (Kugrens 1982, Zuccarello and
359 West 1994). In their study on *L. pacifica* host specificity, Zuccarello and West (1994)
360 found that parasite spores isolated from one host genus were unable to infect members of
361 the other host genus. Kugrens initially designated the *Leachiella pacifica* type specimen
362 as a parasite infecting *Polysiphonia* spp. from Cattle Point, Friday Harbor, WA, USA

(Kugrens 1982). Based upon this, here we are considering the specimen collected on *Polysiphonia hendryi* from Friday Harbor, WA, USA as true *Leachiella pacifica*. Subsequently, parasites infecting *Polysiphonia paniculata* (and several other species) have also been attributed to *L. pacifica* (Zuccarello and West 1994, Zuccarello et al. 2004). Molecular data here identify two distinct species of *Leachiella* that infect separate hosts but are otherwise difficult to distinguish. Due to the highly reduced morphology of these parasites, it seems likely that host identity may be the easiest way to identify parasites in the genus *Leachiella* to species.

Some parasitic red algae are reportedly capable of infecting a range of hosts [e.g., Asterocolax – see Goff, Ashen, and Moon (1997)], however molecular data are contradicting that notion. Rather, it appears that red algal parasite species have higher host specificity than previously believed, and instead, we have underestimated the amount of parasite diversity as a result of their reduced morphology. Morphology based studies report *Leachiella pacifica* infecting at least seven unique host species (Zuccarello and West 1994), further molecular analyses will surely uncover additional species in the genus *Leachiella*.

Parasite Plastids

The link between plastid origin in red algal parasites and their evolutionary relationship to their hosts may be central to our understanding of red algal parasite evolution. The *Harveyella mirabilis* plastid genome (Figure 2) represents the second red algal parasite demonstrated to retain a reduced native plastid. Similar to *Choreocolax polysiphoniae*, the *H. mirabilis* plastid genome remains conserved for functions including

386 amino acid, fatty acid, and protein biosynthesis, but has lost genes involved in building
387 the light harvesting apparatus, photosystems I and II, and other photosynthesis related
388 genes. Identifying a monophyletic clade of parasites (Figure 1) in which the two earliest
389 branching members identified so far retain plastids, strongly suggests that plastids are
390 also retained in species of *Leachiella*. Preliminary data generated in the Lane lab supports
391 that hypothesis, though work remains to completely sequence plastids from *Leachiella*
392 species.

393 The *H. mirabilis* plastid gene order, with the exception of the missing
394 photosynthesis genes, is conserved when compared to plastids of free-living
395 Rhodomelaceae species (Figure 3). In contrast, we find that the *C. polysiphoniae* plastid
396 has undergone greater gene loss than the *H. mirabilis* plastid, and a substantial amount of
397 genome reorganization (Figure 4). *Harveyella mirabilis* retains 13 genes that have been
398 lost from the *C. polysiphoniae* plastid (Table 2); *arg*B and *car*A, which are involved in
399 arginine biosynthesis processes, *rpo*Z, which promotes RNA polymerase assembly, nine
400 ribosomal protein genes, and an uncharacterized hypothetical protein. In contrast, *C.*
401 *polysiphoniae* retains copies of *dna*B, which is involved in DNA replication, and *fab*H,
402 which is involved in fatty acid biosynthesis, both of which have been lost in *H. mirabilis*.
403 Analysis of additional plastid genomes in this clade will provide greater insights into
404 patterns of plastid genome evolution in red algal parasites.

405 Two hypotheses have been proposed for the origin of red algal parasites. Setchell
406 (1918) suggested that a mutation in a spore causes parasites to arise sympatrically, while
407 Sturch (1926) postulated that parasitic red algae start out as epiphytes that over time
408 become endophytes and increasingly rely on the host for nutrition. It seems plausible that

409 Setchell's origin hypothesis could explain the rise of the so-called adelphoparasites,
410 which comprise the majority of known red algal parasite biodiversity. By evolving from
411 their hosts, adelphoparasites could more easily incorporate a copy of a genetically similar
412 plastid as their own. Replacing a functionally reduced or non-functional native plastid
413 with a host-derived plastid seems to be an easy mechanism for survival. Unfortunately,
414 reliance on the host plastid also starts the newly evolved parasite down a path towards
415 inevitable extinction. In order for a non-photosynthetic parasite to survive it still must
416 retain compatibility for other plastid functions including amino acid and fatty acid
417 biosynthesis. Because the host-derived plastid is newly acquired during each new
418 infection, the host plastid experiences one set of evolutionary pressures while the parasite
419 evolves and accumulates mutations of its own. Eventually the parasite will inevitably lose
420 the ability to communicate with the host plastid as the parasite and host increasingly
421 become genetically distinct. This leaves the parasite with two possibilities for survival,
422 either to find another host with a compatible plastid or go extinct. Alternatively, the
423 success of an alloparasite may be explained by Sturch's hypothesis. By evolving from a
424 closely related epiphyte that is able to create secondary pit connections, the parasite may
425 retain its own plastid and therefore enable its longevity and the opportunity to even
426 speciate as the parasite adapts to new hosts. Therefore, what were previously viewed as
427 competing hypotheses to explain the evolution of red algal parasites, may each explain
428 how different types of parasites arise.

429 Although organisms have transitioned from photosynthesis to other methods of
430 nutrient acquisition numerous times across the tree of life, support for total plastid loss is
431 rare (Gornik et al. 2015) with most cases resulting in highly reduced plastids retained for

core functions (e.g. isoprenoid and fatty acid biosynthesis) (Wolfe et al. 1992, Knauf and Hachtel 2002, de Koning and Keeling 2006) or in extreme cases the retention of the apicoplast in Apicomplexans . The non-photosynthetic red algal parasites traditionally considered adelphoparasites, including *Gracilariophila oryzoides* and *Gardneriella tuberifera*, still require a copy of the host plastid for functions that, to date, remain undetermined. Interestingly, in the *Gracilariophila oryzoides* genome and transcriptome, photosynthesis related nuclear-encoded plastid-targeted genes remain conserved while those same genes are absent from the *Choreocolax polysiphoniae* transcriptome (Salomaki and Lane, unpublished).

Data presented here further highlight the need to abandon the notion that taxonomy defines red algal parasites. By placing the family Choreocolacaceae in synonymy with the Rhodomelaceae we are making steps to remove the artificial appearance of parasites infecting hosts in different families. Furthermore, we recognize that a major distinction between types of red algal parasites is the origin of the parasite plastids. While some parasites, including *Harveyella mirabilis* and *Choreocolax polysiphoniae*, retain a native plastid that evolves in concert with the parasite, others including *Gracilariophila oryzoides* and *Gardneriella tuberifera*, incorporate a copy of a host plastid when packaging spores, but will inevitably evolve and become incompatible with the host plastid and blink out of existence. We predict that investigations of parasites that have traditionally referred to as alloparasites, like *Holmsella* and *Pterocladiophila,* will also provide evidence of plastid retention and monophyletic clades of parasites.

ACKNOWLEDGEMENTS

455 The authors thank Amanda Savoie for help identifying parasite sequence data in the UNB
456 database, Jillian Freese for collecting specimens of *Harveyella mirabilis* and *Leachiella*
457 *pacifica,* and Kristina Terpis for her support in the lab. Funding for this work was
458 provided to by a Phycological Society of America Grant-in-Aid of Research to ES and by
459 grant #1257472 from the National Science Foundation awarded to CL. This research is
460 based in part upon work conducted using the Rhode Island Genomics and Sequencing
461 Center which is supported in part by the National Science Foundation (MRI Grant No.
462 DBI-0215393 and EPSCoR Grant Nos. 0554548 & EPS-1004057), the US Department of
463 Agriculture (Grant Nos. 2002-34438-12688 and 2003-34438-13111), and the University
464 of Rhode Island.

465

466 **LITERATURE CITED**

467 Blouin, N. & Lane, C. 2015. Red algae provide fertile ground for exploring parasite
468 evolution. *Perspect. Phycol.* 3:11–9.
469 Broadwater, S.T. & Lapointe, E.A. 1997. Parasitic interactions and vegetative
470 ultrastructure of Choreonema thuretii (Corallinales, Rhodophyta). *J. Phycol.*
471 33:396–407.
472 Darling, A.C.E., Mau, B., Blattner, F.R. & Perna, N.T. 2004. Mauve : Multiple alignment
473 of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–403.
474 de Koning, A.P. & Keeling, P.J. 2006. The complete plastid genome sequence of the
475 parasitic green alga Helicosporidium sp. is highly reduced and structured. *BMC Biol.*
476 4.
477 Feldmann, G. & Feldmann, J. 1958. Recherches sur quelques Floridees parasites. *Rev.*

Gen. Bot. 65:49–128.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C. et al. 2015. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44:D279–85.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L. et al. 2010. The Pfam protein families database. *Nucleic Acids Res.* 38:D211-22.

Goff, L.J. 1982. The biology of parasitic red algae. *Prog. Phycol. Res.* 1:289–369.

Goff, L.J., Ashen, J. & Moon, D. 1997. The evolution of parasites from their hosts: A case study in the parasitic red alga. *Evolution.* 51:1068–78.

Goff, L.J. & Cole, K. 1975. The biology of Harveyella mirabilis (Cryptonemiales, Rhodophyceae). II . Carposporophyte development as related to the taxonomic affiliation of the parasitic red alga, Harveyella mirabilis. *Phycologia*. 14:227–38.

Goff, L.J. & Coleman, A.W. 1985. The role of secondary pit connections in red algal parasitism. *J. Phycol.* 21:483–508.

Goff, L.J. & Coleman, A.W. 1995. Fate of parasite and host organelle DNA during cellular transformation of red algae by their parasites. *Plant Cell*. 7:1899–911.

Goff, L.J., Moon, D.A., Nyvall, P., Stache, B., Mangin, K. & Zuccarello, G. 1996. The evolution of parasitism in the red algae: Molecular comparisons of adelphoparasites and their hosts. *J. Phycol.* 32:297–312.

Gornik, S.G., Febrimarsa, Cassin, A.M., MacRae, J.I., Ramaprasad, A., Rchiad, Z., McConville, M.J. et al. 2015. Endosymbiosis undone by stepwise elimination of the plastid in a parasitic dinoflagellate. *Proc. Natl. Acad. Sci.* 112:201423400.

Happe, T. & Naber, J.D. 1993. Isolation, characterization and N-terminal amino acid

501 sequence of hydrogenase from the green alga Chlamydomonas reinhardtii. *Eur. J.*
502 *Biochem.* 214:475–81.
503 Jacobs, J., Pudollek, S., Hemschemeier, A. & Happe, T. 2009. A novel, anaerobically
504 induced ferredoxin in Chlamydomonas reinhardtii. *FEBS Lett.* 583:325–9.
505 Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. 2016. KEGG as a
506 reference resource for gene and protein annotation. *Nucleic Acids Res.* 44:D457–62.
507 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S.
508 et al. 2012. Geneious Basic: an integrated and extendable desktop software platform
509 for the organization and analysis of sequence data. *Bioinformatics*. 28:1647–9.
510 Knauf, U. & Hachtel, W. 2002. The genes encoding subunits of ATP synthase are
511 conserved in the reduced plastid genome of the heterotrophic alga Prototheca
512 wickerhamii. *Mol. Genet. Genomics*. 267:492–7.
513 Kugrens, P. 1982. Leachiella pacifica, Gen. et Sp. nov., a new parasitic red alga rom
514 Washington and California. *Am. J. Bot.* 123:306–19.
515 Kurihara, A., Abe, T., Tani, M. & Sherwood, A.R. 2010. Molecular phylogeny and
516 evolution of red algal parasites: A case study of Benzaitenia, Janczewskia, and
517 Ululania (Ceramiales). *J. Phycol.* 46:580–90.
518 Le Gall, L. & Saunders, G.W. 2010. DNA barcoding is a powerful tool to uncover algal
519 diversity: A case study of the phyllophoraceae (Gigartinales, Rhodophyta) in the
520 Canadian flora1. *J. Phycol.* 46:374–89.
521 Preuss, M., Nelson, W.A. & Zuccarello, G.C. 2017. Red algal parasites: a synopsis of
522 described species, their hosts, distinguishing characters and areas for continued
523 research. *Bot. Mar.* 60:1–13.

524 Reinsch, P.F. 1875. Contributiones ad Algologiam et Fungologiam. Lipsiae, Nuremberg.
525 61 pp.
526 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget,
527 B. et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model
528 choice across a large model space. *Syst. Biol.* 61:539–42.
529 Salomaki, E.D. & Lane, C.E. 2014. Are all red algal parasites cut from the same cloth?
530 *Acta Soc. Bot. Pol.* 83:369–75.
531 Salomaki, E.D. & Lane, C.E. 2016. Red Algal Mitochondrial Genomes are More
532 Complete than Previously Reported. *Genome Biol. Evol.* 48:evw267.
533 Salomaki, E.D., Nickles, K.R. & Lane, C.E. 2015. The ghost plastid of Choroeocolax
534 polysiphoniae. *J. Phycol.* 51:217–21.
535 Saunders, G.W. 1993. Gel purification of red algal genomic DNA: An inexpensive and
536 rapid method for the isolation of polymerase chain reaction-friendly DNA. *J.*
537 *Phycol.* 29:251–4.
538 Saunders, G.W. 2005. Applying DNA barcoding to red macroalgae: a preliminary
539 appraisal holds promise for future applications. *Philos. Trans. R. Soc. Lond. B. Biol.*
540 *Sci.* 360:1879–88.
541 Saunders, G.W. 2014. Long distance kelp rafting impacts seaweed biogeography in the
542 Northeast Pacific: the kelp conveyor hypothesis. *J. Phycol.* 50:968–74.
543 Saunders, G.W. & Moore, T.E. 2013. Refinements for the amplification and sequencing
544 of red algal DNA barcode and RedToL phylogenetic markers: a summary of current
545 primers, profiles and strategies. *Algae*. 28:31–43.
546 Schattner, P., Brooks, A.N. & Lowe, T.M. 2005. The tRNAscan-SE, snoscan and

547 snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.*
548 33:686–9.
549 Setchell, W.A. 1914. Parasitic Florideae I. *Univ. Calif. Publ. Bot.* 6:1–34.
550 Setchell, W.A. 1918. Parasitism Among the Red Algae. *Proc. Am. Philisophical Soc.*
551 57:155–72.
552 Setchell, W.A. 1923. Parasitic Florideae II. *Univ. Calif. Publ. Bot.* 10:393–401.
553 Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-
554 analysis of large phylogenies. *Bioinformatics*. 30:1312–3.
555 Sturch, H.H. 1926. Choreocolax Polysiphoniae, Reinsch. *Ann. Bot.* XL.
556 The UniProt Consortium. 2017. UniProt: the Universal Protein knowledgebase. *Nucleic*
557 *Acids Res.* 45:D158–69.
558 Vaidya, G., Lohman, D.J. & Meier, R. 2011. SequenceMatrix: Concatenation software
559 for the fast assembly of multi-gene datasets with character set and codon
560 information. *Cladistics*. 27:171–80.
561 Wetherbee, R. & Quirk, H. 1982. The Fine Structure and Cytology of the Association
562 Between the Parasitic Red Alga Holmsella australis and Its Red Algal Host
563 Gracilaria furcellata. *Protoplasma*. 165:153–65.
564 Wolfe, K.H., Morden, C.W. & Palmer, J.D. 1992. Function and evolution of a minimal
565 plastid genome from a nonphotosynthetic parasitic plant. *Proc. Natl. Acad. Sci. U. S.*
566 *A.* 89:10648–52.
567 Zuccarello, G.C., Moon, D. & Goff, L.J. 2004. A phylogenetic study of parasitic genera
568 placed in the family Choreocolacaceae (Rhodophyta). *J. Phycol.* 40:937–45.

569 Zuccarello, G.C. & West, J.A. 1994. Genus and race specificity in the red algal parasite

570 Leachiella pacifica (Choreocolacaceae, Rhodophyta). *Phycologia*. 33:213–8.

571

572

573 Table 1. GenBank Sequence accession numbers for taxa used in phylogenetic analyses
574 (Figure 2).
575

	18S	**28S**	**COI-5P**	***rbc*L**
Amansia fimbrifolia	HM582913	HQ422079	HM582889	-
Amansia glomerata	HM582909	HQ422225	HQ422913	-
Bostrychia arbuscula	-	AY920894	KM502795	AY920845
Bostrychia intricata	-	KM502850	KM502801	AY920846
Bostrychia vaga	-	KM502842	KM502791	KM502817
Choreocolax polysiphoniae	This Study	This Study	This Study	-
Harveyella mirabilis	This Study	This Study	This Study	-
Heterodasya mucronata	-	-	KC567673	KF367797
Leachiella pacifica	This Study	This Study	This Study	-
Leachiella sp. GWS021225	This Study	This Study	KM254767	-
Neorhodomela larix	AY617140	-	KM254241	GQ252553
Neorhodomela oregona	-	-	KM254304	GQ252556
Odonthalia dentata	-	JX572172	JX571960	KU564463
Odonthalia floccosa	AY617141	-	KM254276	GQ252492
Osmundaria obtusiloba	-	HQ422003	HQ422818	-
Rhodomela confervoides	AY617145	KX145642	KX258842	KX146197
Rhodomela lycopodioides	-	This Study	HM916516	KU564489

576
577

578 Table 2. *Harveyella mirabilis* plastid genes and their function, which have been lost from
579 the *Choreocolax polysiphoniae* plastid genome.
580

Gene Name	**Biological Process and Molecular Function**
*arg*B	Amino acid biosynthesis; ATP binding, acetylglutamate kinase activity
*car*A	Pyrimidine metabolism; ATP binding, carbamoyl-phosphate synthase (glutamine-hydrolyzing) activity
*inf*B	Translation; Translation initiation, GTP binding
*rpl*9	Translation; structural constituent of the 50S ribosome
*rpl*17	Translation; structural constituent of the 50S ribosome
*rpl*24	Translation; structural constituent of the 50S ribosome
*rpl*32	Translation; structural constituent of the 50S ribosome
*rpl*33	Translation; structural constituent of the 50S ribosome
*rpl*34	Translation; structural constituent of the 50S ribosome
*rpo*Z	Transcription; DNA binding
*rps*18	Translation; structural constituent of the 30S ribosome
*rps*20	Translation; structural constituent of the 30S ribosome
*ycf*21	Uncharacterized hypothetical protein

581

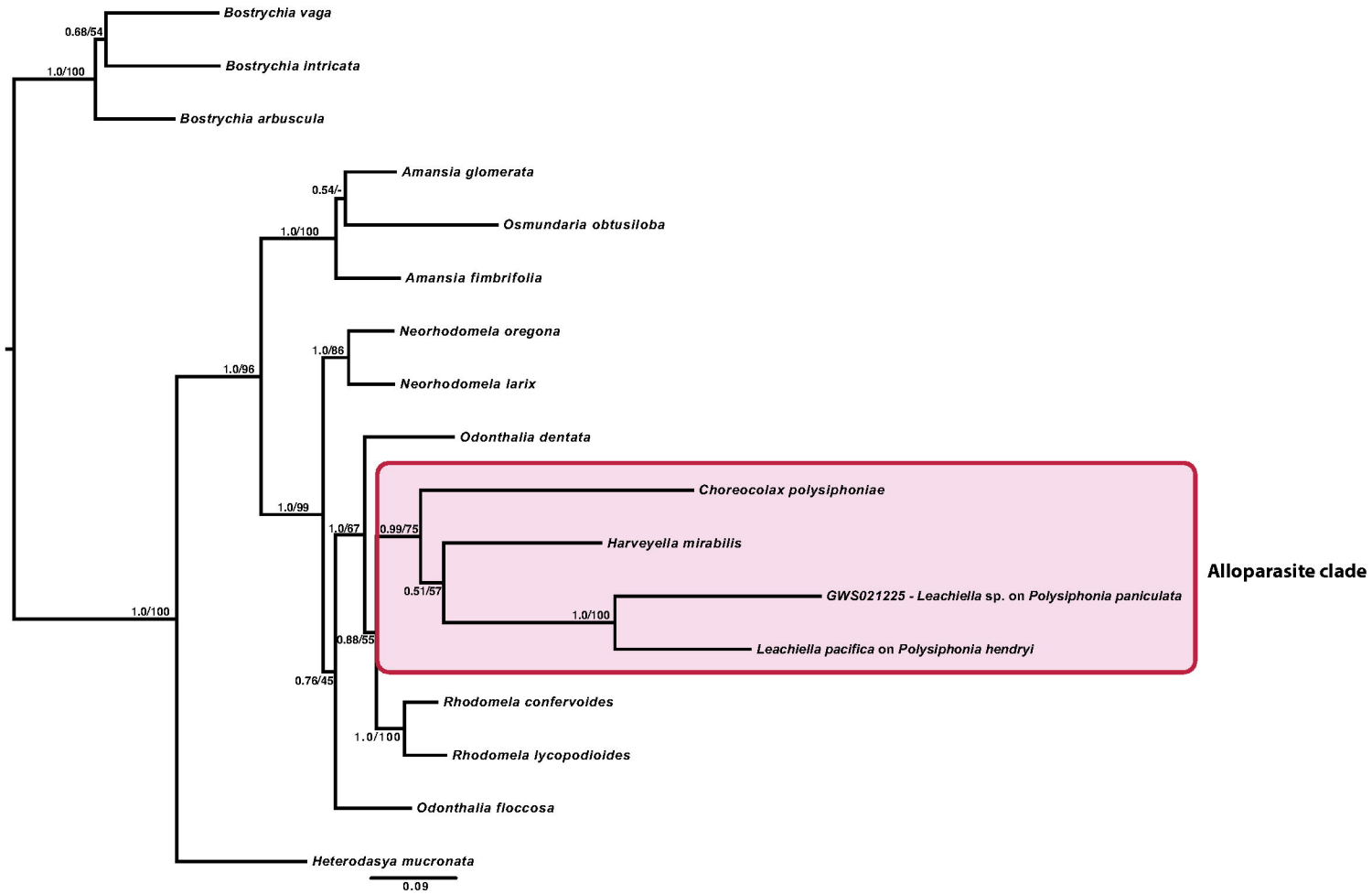

Figure 1. Maximum likelihood phylogeny based on concatenated 18S, 28S, COI-5P, and rbcL data as shown in Table 1, of the parasites *Choreocolax polysiphoniae, Harveyella mirabilis, Leachiella pacifica*, an undescribed species of *Leachiella* found on *Polysiphonia paniculata*, and thirteen close relatives. This phylogenetic analysis supports a monophyletic clade of parasites arising within the Rhodomelaceae. Support values shown as Bayesian posterior probability/maximum likelihood bootstrap.

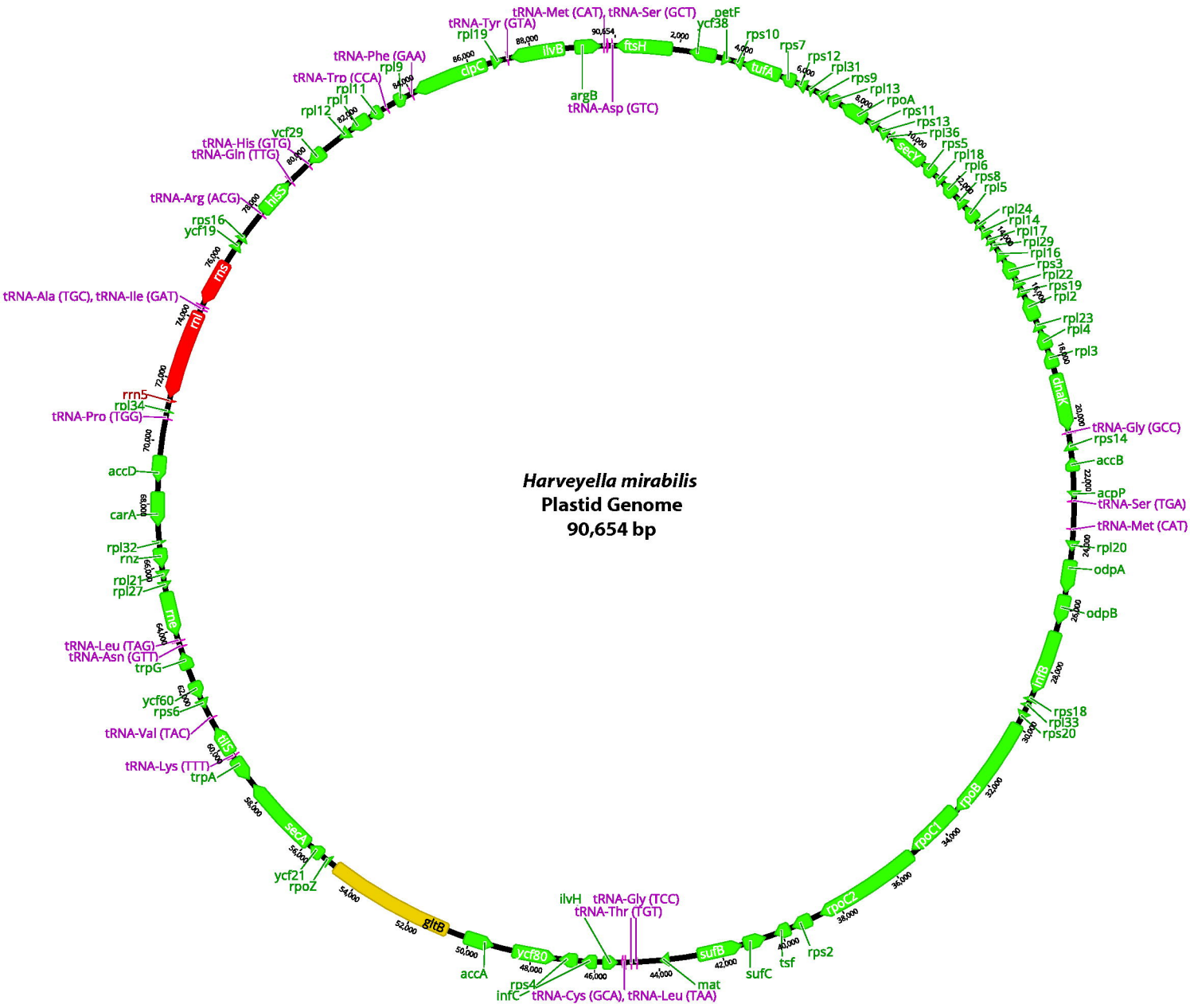

Figure 2. The plastid genome of the parasitic red alga, *Harveyella mirabilis* is 90,654 basepairs and contains 84 protein coding genes (Green), the 5S, 16S, and 23S rRNAs (Red), and 23 tRNAs (Pink). All genes involved with photosynthetic functions, except *pet*F, have been lost. The *fts*H gene is truncated but may still be transcribed, however *glt*B is a non-functional pseudogene (Yellow).

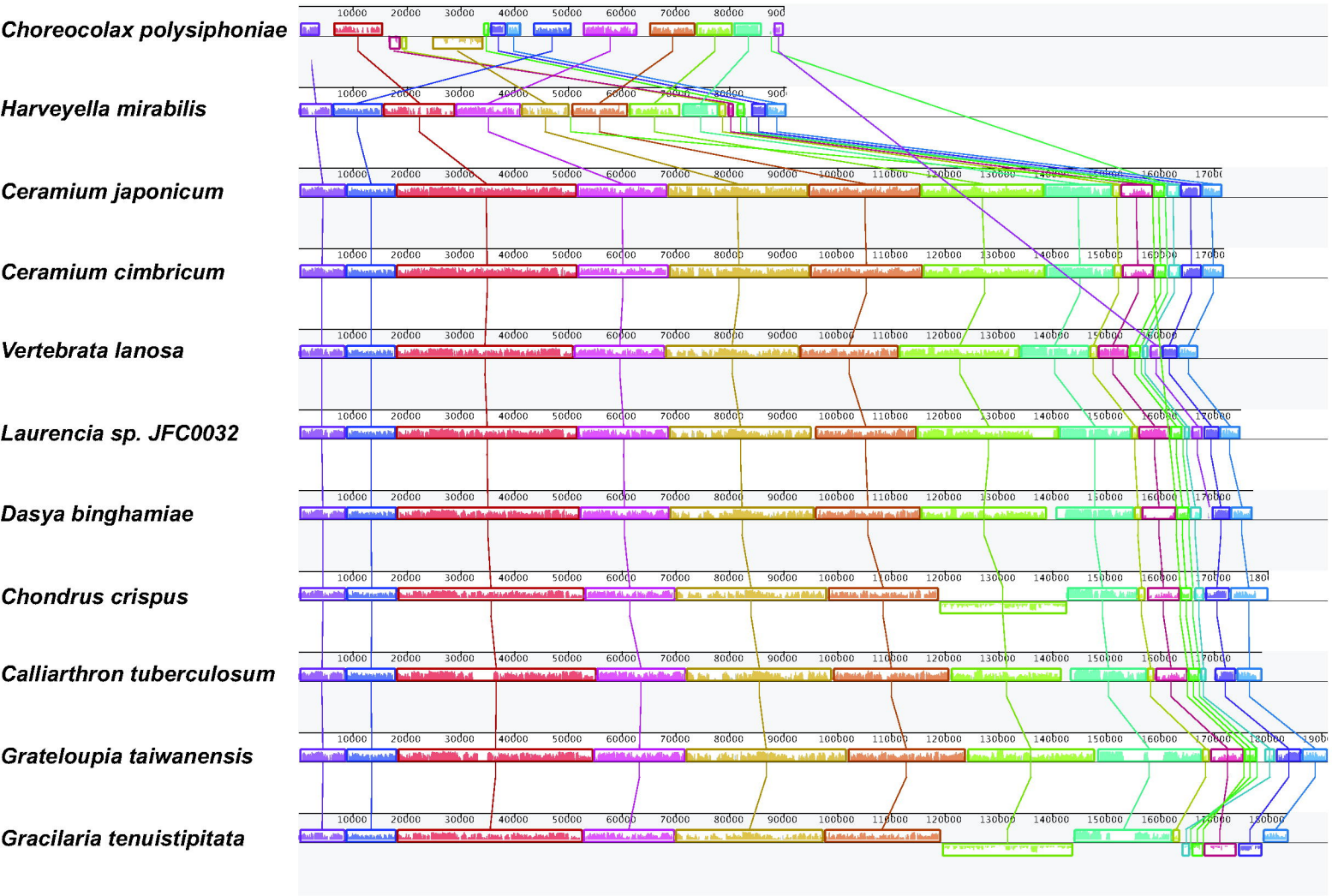

Figure 3. Mauve alignment of the parasites *Choreocolax polysiphoniae* (top) and *Harveyella mirabilis* (second from top) with all published Rhodomelaceae plastid genomes as well as those from select representatives of other Florideophyceae families (*Dasya binghamiae*, Dasyaceae; *Chondrus crispus*, Gigartinaceae; *Calliarthron tuberculosum*, Corallinaceae; *Grateloupia taiwanensis*, Halymeniaceae; *Gracilaria tenuistipitata*, Gracilariaceae). This alignment identifies 16 locally collinear blocks (LCBs) among the selected plastid genomes and demonstrates that even with the loss of photosynthesis genes overall synteny is shared between the parasite *H. mirabilis* and other plastid genomes, while *C. polysiphoniae* has undergone several genome rearrangements.

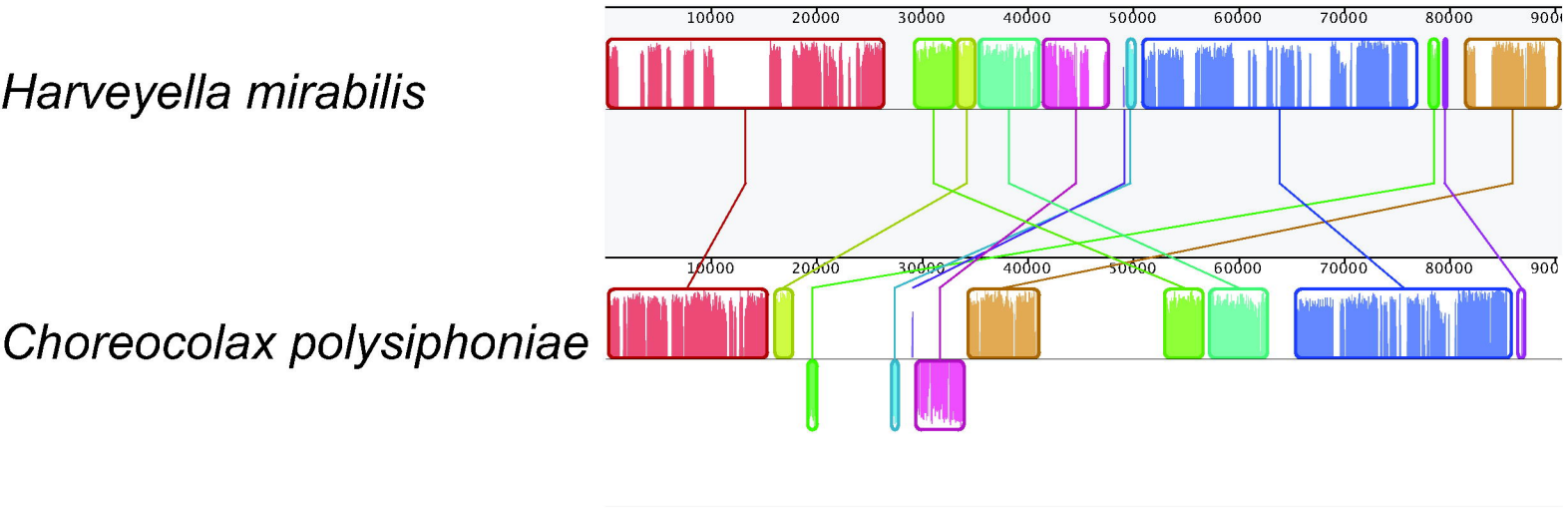

Figure 4. Mauve alignment of the plastid genomes from the parasites *Harveyella mirabilis* (top) and *Choreocolax polysiphoniae* (bottom). This alignment identifies 11 locally collinear blocks (LCBs) among these parasite plastid genomes, highlighting the high level of genome fragmentation and rearrangements evident in the two parasite plastids