Title: Transposable Element Diversification and the Evolution of Peltigerales Lichen Symbionts

Authors: Ellen S. Cameron[1,2,*], Benjamin J.M. Tremblay[3], Rebecca Yahr[4], Mark L. Blaxter[2], Robert D. Finn[1,*]

[1] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge CB10 1SD, United Kingdom

[2] Tree of Life, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom

[3] The Sainsbury Laboratory, University of East Anglia, Norwich Research Park, Colney Lane, Norwich, NR4 7UH, UK

[4] Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, UK

[*]Corresponding authors:

ellenscameron@outlook.com (Ellen S. Cameron)

rdf@ebi.ac.uk (Robert D. Finn)

Keywords

Cyanolichen, metagenomics, genome evolution, symbiosis, transposable element

22 # Table of Contents

23 **Abstract** **4**

24 **1. Introduction** **5**

25 **2. Methods** **6**

26 **2.1 Sample collection, nucleic acid extraction & sequencing** **6**

27 **2.2 Generation of metagenomic assembled genomes (MAGs)** **7**

28 **2.3 Exploration of cyanobacterial symbiont genomes** **8**

29 **2.4 Annotation and characterization of eukaryotic symbionts** **8**

30 **2.5 Characterizing transposable element content** **10**

31 **3. Results** **12**

32 **3.1 Cyanolichens host diverse bacterial and eukaryotic microbial**
33 **communities** **12**

34 Figure 1: Organismal composition of 11 different species of cyanolichens
35 determined using long-read metagenomic approaches 14

36 **3.2 Alternative vanadium-dependent nitrogen-fixation pathways observed in**
37 **Nostoc symbionts** **15**

38 Figure 2: Diversity, distribution and genomic exploration of cyanobacterial
39 metagenomic assembled genomes (MAGs). 17

40 **3.3 Transposable elements drive genome expansion in chlorophyte**
41 **symbiont** **17**

42 Figure 3: Identification and genomic characteristics of the chlorophyte
43 photobiont in Ricasolia virens and Lobaria pulmonaria. 18

44 **3.4 Metagenomic recovery of largest Lecanormycetes genomes to date** **21**

45 Figure 5: Genome size and repeat content of lichenized and non-lichenized
46 fungi.. 21

47 Figure 6: Distribution and expression of genes and TEs in fungal genomes. ... 23

48 **3.5 Cellular stress response genes enriched near transposable elements** .. **24**

49 Figure 7: Expression and functional enrichment of genes in relation to nearest
50 transposable element in fungal metagenomic assembled genomes (MAGs). .. 25
51 Figure 8: Enriched GO terms in Peltigerales fungal metagenomic assembled
52 genomes (MAGs)..... 26
53 **4. Discussion..... 26**
54 **Acknowledgements 29**
55 **Data and code availability 30**
56 **5. References 30**
57
58
59

Abstract

Lichens are composite organisms formed through the symbiotic association between fungi, algae and/or bacteria. Multiple independent origins of the lichenized lifestyle have been reported in both fungal and algal lineages, but the molecular mechanisms and evolution underpinning these symbiotic relationships remain largely unknown. In this study, we performed long-read metagenomic sequencing on 11 Peltigerales lichen species to characterize the genomic content of lichen symbionts via metagenome assembled genomes (MAGs). Peltigerales genomes generated in this work represent the largest Lecanoromycetes genome sequenced to date, driven by high transposable element content. Transposable elements (TEs) are known to drive genome evolution in other symbioses but have been underexplored in lichen symbionts due technological limitations. Transcriptomics revealed that many genes associated with adaptations to the lichenized lifestyle are associated with TEs suggesting that they may play a key role in the evolution of lichenization.

1. Introduction

Lichens have traditionally been viewed as the symbiotic association between a fungal partner and photosynthetic algal partner[1]. It is estimated that there are over 20,000 species of lichenized fungi[2], with most forming symbiotic associations with chlorophyte algae. However, some lichens are formed through the symbiosis between a fungal partner and a cyanobacterium ('cyanolichens'), and in other cases both chlorophyte and cyanobacterial partners are involved ('tripartite cyanolichens')[3,4]. Multiple independent origins of lichenization are inferred in diverse fungal lineages[5,6], and in cyanobacterial[7] and chlorophyte photobionts[8]. More recently, the diversity of what constitutes a lichen has expanded with metagenomic surveys revealing complex multispecies assemblages beyond primary fungal and algal partners, consisting of diverse bacterial communities[9,10], additional fungi[11,12], and viruses[13]. However, fundamental questions of the function and evolution for the two primary symbionts of the lichen, fungi and algae, remain unanswered[1,14]. Long-read metagenomic sequencing provides new avenues to explore these questions further through the assembly of high-quality metagenomic assembled genomes (MAGs), which can be used to explore the genomic characteristics of symbionts suggested to drive the evolution in other fungal interactions.

Transposable elements (TEs), also known as transposons or mobile elements, can move within a genome via transposition and often represent the majority of repetitive sequences found in eukaryotic genomes[15,16]. In fungi, TEs may contribute to key roles in adaptation to symbiotic or pathogenic lifestyles[17,18] with TEs previously being found to be involved with genes associated with symbiosis in ectomycorrhizal fungi[19] or virulence factors in pathogenic fungi[18,20]. Proliferation of TEs can result in expansions in genome size and the generation of new mutations, but mechanisms exist to repress TE activity including epigenetic silencing and loss through degradation[21]. In rare cases, TE insertions may be beneficial to their hosts, leading to domestication and retention in the host genome[21] or contribute to genetic variability driving the coevolution between symbionts[22]. The proportion of eukaryotic genomes made up of TEs ranges from only 3% in yeast to upwards of 80% in some plants[17]. However, in some fungi with host-associated lifestyles, TEs contribute to

106 large proportions of the total genome size as observed in arbuscular mycorrhizae
107 (47%)[18] or the fungal pathogen, *Austropuccinia psidii* (>90%) which has the largest
108 fungal genome sequenced to date[19]. Despite TE content being identified as a
109 potential driver of coevolution in plant-associated fungi with diverse lifestyles ranging
110 from pathogenic to mutualism, few papers have addressed the role of TEs in other
111 fungal symbioses including lichens[23,24]. The ability to explore the role of TEs rely on
112 high-quality, highly contiguous genomic assemblies which may not be consistently
113 achieved with short-read shotgun metagenomics[9].
114
115 In this study we used long-read metagenomics and short-read transcriptomics to
116 characterize all the organismal constituents and their associated genomic content,
117 including TE content, for 11 different species of cyanolichens. We generated highly
118 contiguous metagenomic assembled genomes (MAGs) for primary fungal symbionts
119 and photobionts of cyanolichens, including the first genome assemblies for ten of the
120 lichenized fungal species. Using these MAGs, we demonstrated that TE content may
121 contribute to the evolution in Peltigerales fungi by (1) comprising a large portion of
122 the genomic content and (2) being associated with highly expressed genes
123 associated with stress response and cellular communication.

124 **2. Methods**

125 *2.1 Sample collection, nucleic acid extraction & sequencing*

126 Lichen thalli of 11 cyanolichen species (Table S1) were collected with permission
127 from Glen Creran Woods Site of Special Scientific Interest (Argyll, Scotland) and air
128 dried for preservation for 72 hours. Associated bryophyte material was removed from
129 lichen thallus prior to long-term storage in a -80°C freezer.
130 Thallus tissue for samples used in analysis were prepared for nucleic acid extraction
131 using a CryoPrep for flash frozen tissue pulverisation. The resulting pulverised tissue
132 powder was used for DNA extraction using the protocol proposed by Park et al.
133 (2014)[25], which utilises a potassium chloride extraction buffer with chloroform (Table
134 S2). DNA extracts were assessed for quantity and quality using a NanoDrop
135 spectrophotometer and Qubit fluorometric assay. DNA extracts were sheared using
136 a MegaRuptor at a speed of 130, and SPRI cleaned up (0.88 v/v concentration) with
137 the addition of RNase to degrade any remaining RNA material. Sheared DNA

138 extracts were assessed for suitable fragment length using the Agilent FemtoPulse.
139 Final DNA extracts were submitted for PacBio HiFi sequencing on the Revio platform
140 at Scientific Operations at the Wellcome Sanger Institute.
141
142 Thallus tissue from two individuals or two parts of a single individual of each species
143 used for genomic sequencing were pulverized for RNA extraction (Table S3). RNA
144 extractions were performed following the manufacturer's protocol for TRIzol including
145 a second chloroform wash step prior to isopropanol precipitation to remove phenol
146 contamination. RNA extracts were treated with DNase using the Turbo DNase kit to
147 remove remaining DNA. Resulting RNA extracts were assessed for quantity and
148 quality using an Agilent RNA Screen Tape Station analysis and were submitted for
149 NEB NextERA RNAseq Illumina sequencing on the NovaSeqX platform at Scientific
150 Operations at the Wellcome Sanger Institute.
151
152 *2.2 Generation of metagenomic assembled genomes (MAGs)*

153 PacBio HiFi reads were assembled into longer contiguous sequences using
154 *metaMDBG* (v. 0.3; Table S4)[26]. Reads were mapped back to resulting metagenomic
155 assemblies using *minimap2* (v. 2.23)[27] to determine assembly coverage. Contigs
156 were binned into representative genomes using *CONCOCT* (v. 1.1.0)[28] and
157 *metaBAT2* (v. 2.12.1)[29].
158
159 Eukaryotic bins were identified from the set of bins generated with *metaBAT2* with
160 *EukCC* (v. 2.1.2)[30] . Eukaryotic bins were assessed for completeness using *EukCC*
161 and *BUSCO* (v. 5.7.0; Table S5; Table S6)[31], and merged bins were generated using
162 *EukCC* where applicable.
163
164 To explore the bacterial portion of the lichen metagenome, bins from *CONCOCT* and
165 *metaBAT2* were refined using the *metaWrap* (v. 1.3)[32] bin refinement module.
166 Refined bin sets were dereplicated using *dRep* (v. 3.2.2; Table S7)[33] and
167 taxonomically classified using *GTDBtk* (v. 2.3.0)[34]. Quality scores (QS; completeness
168 - [5 * contamination]) were calculated for bacterial bins using metrics determined by
169 *CheckM* (v. 1.1.3; Table S8)[35]. Bacterial bins with a QS > 50 were retained for gene
170 annotation with *prokka* (v.1.14.6)[36]. Predicted genes were used for functional

171 annotations with *InterproScan* (v. 5.60-92.0)[37]. For phylogenetic tree reconstruction,
172 ModelFinder[38] was used to identify the best-fit model according to the Bayesian
173 Information Criterion (BIC). Subsequently, a phylogenetic tree for bacterial MAGs
174 was generated using alignments from *GTDBtk* with *iqtree* (v. 2.2.0)[39] using model
175 LG+F+I+G4 with 1000 bootstraps and visualized with *ggtree* (v. 3.12.0)[40]
176
177 Genomic composition of metagenomes was assessed by mapping reads with
178 *minimap2* to assemblies and calculating a weighted mean depth as a function of the
179 length of the MAG (Table S9) and contig to the number of reads mapped. MAGs
180 were identified as being 'present' in a metagenome if >30% of the MAG was covered
181 by reads (Table S10; Table S11).
182
183 *2.3 Exploration of cyanobacterial symbiont genomes*

184 Synteny analysis of cyanobacterial MAGs was performed using *jcvi*[41] to identify
185 genomic arrangements between different detected *Nostoc* species. KEGG orthologs
186 identified with KofamScan (v. 1.3.0)[42] annotations corresponding to nitrogen fixation
187 pathways were identified in cyanobacterial MAGs to explore the occurrence of iron-
188 molybdenum dependent and vanadium dependent nitrogen fixation pathways in
189 cyanobacterial symbionts. Cyanobacterial MAGs were placed into a phylogenetic
190 tree with other reference genomes (Table S12) to explore habitat associations.
191 Cyanobacterial MAGs and reference Nostocaceae and Leptolyngbyaceae genomes
192 were dereplicated using *dRep* (Table S13). Alignments were generated for the
193 dereplicated set of cyanobacterial genomes, including both references and MAGs,
194 using GTDBtk. ModelFinder identified the best-fit model according to the BIC and a
195 phylogenetic tree created with *iqtree* using model LG+F+R8 with 1000 bootstraps
196 and visualized with *iTOL* (v. 7)[43].
197
198 *2.4 Annotation and characterization of eukaryotic symbionts*

199 Genome assemblies for fungi (Lecanoromycetes; Table S14) and green algae
200 (Chlorophyta; Table S15) were downloaded from NCBI in June 2024 and retained for
201 analysis if the N50 was greater than 1Mb to ensure the retention of high-quality
202 genome assemblies. Phylogenetic trees were generated for fungal and algal
203 lineages using BUSCO phylogenomics

204 (https://github.com/jamiemcg/BUSCO_phylogenomics) in which BUSCO groups
205 shared amongst 90% of genomes are concatenated and aligned for placement.
206 ModelFinder was used to identify the best model according to BIC. A phylogenetic
207 tree was constructed for the Lecanoromycetes assemblies using the JTT+F+I+G4
208 model with 1000 bootstraps. For the Chlorophyta assemblies, the LG+F+I+G4 model
209 was used for phylogenetic tree construction with 1000 bootstraps in *iqtree*.
210 Telomeres were identified in eukaryotic MAGs using a script from Hiltunen *et al.*, with
211 a query search of TAA[C]+ for fungal genomes and CCCATTT for algal genomes as
212 described in Tagirdzhanova *et al.*[44,45] GC content was calculated per contig with a
213 sliding window size of 10,000 using *universalmotif* (v. 1.22.3)[46].
214
215 To further explore the taxonomic identity of the chlorophyte photobiont, 18S rRNA
216 gene sequences were identified in the recovered chlorophyte photobiont MAG with
217 *barrnap* (v. 0.9)[47] and placed phylogenetically with other 18S rRNA gene sequences
218 from other lichenized and non-lichenized chlorophyte algae (Table S16) using the
219 TNe+I+G4 model as identified by ModelFinder in *iqtree*.
220
221 Putative organellar genomes for the chlorophyte photobiont species were identified
222 using *tiara* (v. 1.0.3)[48] classifications and confirmed based on:(i) size;(ii) circularity;
223 and (iii) gene annotation. Chloroplast genomes for chlorophyte photobionts were
224 assessed for completeness using quality scoring algorithms present in *plastiC*
225 (v.0.1.2)[49]. Mitochondrial and chloroplast genomes were annotated and visualized
226 with *OG-DRAW* (v. 1.3.1)[50].
227
228 Initial protein predictions were performed on eukaryotic MAGs using *metaEuk* (v. 7-
229 bba0d80)[51] with a reference database corresponding to respective lineages of
230 Trebouxiophyceae for chlorophyte photobionts and Lecanoromycetes for lichenized
231 fungi. These protein predictions were used alongside MAGs to characterize splicing
232 characteristics and were used to generate an index with *STAR* (v. 2.7.11b)[52] for each
233 individual species. Raw RNAseq reads were trimmed using *fastp* (v. 0.23.4)[53] and
234 mapped to MAGs using *STAR. stringtie* (v. 2.2.3)[54] was then used to predict
235 transcripts based on this mapping, using BAM files produced by *STAR* split into
236 forward and reverse strands using *samtools* (v. 1.21)[55] to improve the annotation of
237 overlapping genes. Final transcript sequences were generated with *gffread* (v.

238 0.12.7)[56] and were used for coding sequence (CDS) generation with *Transdecoder*
239 (v. 5.7.1)[57] including homology searches conducted with *hmmer* (v. 3.1b2) and
240 *diamond* (v. 2.0.14)[58]. Gene expression was quantified using *salmon* (v. 1.10.3)[59].
241
242 For samples with more than one eukaryotic MAG present (including tripartite
243 cyanolichens, and those with additional fungi), eukaryotic bins and GTF files for each
244 detected species were merged to generate a combined species index with *STAR* for
245 mapping, and subsequent steps were followed as highlighted above.
246
247 The gene annotations as generated above were functionally annotated with
248 *InterProScan*.
249
250 *2.5 Characterizing transposable element content*

251 Repeat content of eukaryotic genomes was characterized using *EarlGrey* (v. 5.1)[60].
252 Transposable elements less than 100 bp in size were discarded from downstream
253 analyses. Overall repeat content (%) of the genome was determined based on
254 coverage and length of identified transposable elements. The co-occurence of
255 transposable elements and genes were visualized using *Circlize*[61]. Kimura distances,
256 as determined by *EarlGrey*, were used to characterize the TE activity to identify TE
257 acquisition events in the repeat landscape based on similarity to consensus
258 sequences.
259
260 Transposable element position was used in reference to gene positions to determine
261 genes located within or in close proximity to transposable elements. Distances of
262 genes to transposable elements were classified into categorical bins as follows: 0bp
263 (e.g., gene inside TE, fragments of TEs inside genes), 1-1,000bp, 1,001-5,000bp,
264 5,001-10,000bp, 10,001-20,000bp, >20,000 bp. To identify gene-rich or TE-rich
265 regions, contigs were split into 25 kb windows, sliding every 5 kb. Windows were
266 assigned as gene-rich if they contained over 60% genes, and TE-rich if they
267 contained over 60% TEs. Overlapping windows with identical assignments were
268 merged together.
269

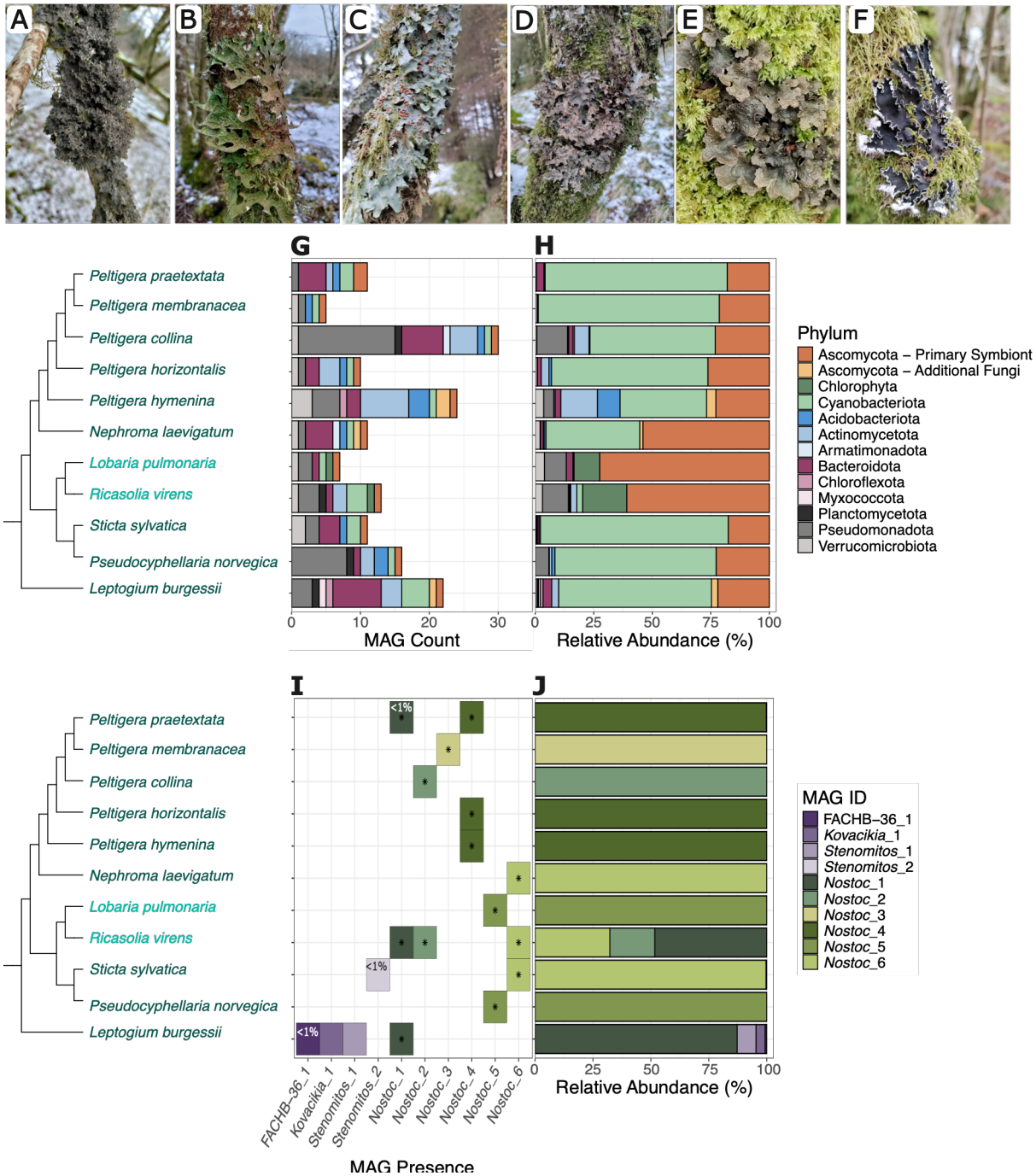
270 Gene expression was explored to identify correlation of highly expressed genes,
271 proximity to transposable elements, and functional relevance. Three approaches
272 were used to examine the relationship between gene expression and TE proximity.
273 First, the average log of the normalized values of transcripts per million (TPM)
274 obtained from *salmon* for each species with the distance between each gene with
275 their nearest TE were plotted as scatter plots. Second, the average log TPM values
276 per species were grouped into the previously determined gene-rich, TE-rich, or
277 mixed categories and plotted as boxplots. Third, the distribution of expressed
278 (average log TPM value greater or equal to 1) and non-expressed (average log TPM
279 value less than 1) gene counts per species per categorical TE distance bins were
280 plotted as stacked bar plots.
281
282 Functional enrichment of genes grouped by categorical TE distance bins was
283 performed using the *topGO* (v. 2.56.0)[62] using Gene Ontology (GO) Biological
284 Processes (BP) annotations determined by *InterProScan* to calculate enrichment
285 within each bin. Selected enriched GO terms (p-value < 0.05) with biological
286 relevance were used for visualization. All enriched GO terms are available in
287 supplementary data (Table S20/S21).
288
289 The *regioneR* (v. 1.36.0)[63] package was used to determine whether genes are
290 present closer or further away from TEs in the lichenized fungal genomes than
291 expected by chance by shuffling the locations of the genes in the genomes 1,000
292 times and calculating the average gene to TE distance for each iteration, leading to a
293 final P-value calculation. The same procedure was carried out to determine whether
294 specific GO:BP annotated genes were closer or further away, only instead shuffling
295 the annotations between existing genes with annotated GO terms. Terms closer or
296 further away than the average were semantically enriched and clustered using the
297 *simplifyEnrichment* (v. 1.14.1)[64] package. The *regioneR* package was also used to
298 test whether specific TE families are closer or further away from genes than
299 expected by average, shuffling the family label of TEs (excluding TEs with no known
300 family).
301
302 Candidate TE fusions were identified using TE-specific Pfam domains previously
303 identified in fungi[65]. Transcripts were characterized as TE-derived if all identified

304 Pfam domains within the protein sequence were TE-specific, and TE fusions if the
305 Pfam domains were a mix of TE-specific and non-TE-specific domains.

306 **3. Results**

307 *3.1 Cyanolichens host diverse bacterial and eukaryotic microbial communities*

308 PacBio HiFi metagenomic sequencing was conducted for 11 cyanolichen species
309 (Figure 1A-F) leading to the generation of high-quality metagenomic assemblies
310 (Table S4). Metagenomic assembled genomes (MAGs) for primary fungal,
311 cyanobacterial and chlorophyte symbionts, as well as additional fungi and bacteria
312 were found in the lichen thallus (Table S10/S11). Following genome redundancy
313 removal, a total of 103 bacterial, 15 fungal and 1 chlorophyte MAGs were identified
314 (Figure 1G).
315


A
B
C
D
E
F
G
H
Peltigera praetextata
Peltigera membranacea
Peltigera collina
Peltigera horizontalis
Peltigera hymenina
Nephroma laevigatum
Lobaria pulmonaria
Ricasolia virens
Sticta sylvatica
Pseudocyphellaria norvegica
Leptogium burgessii
0
10
20
30
MAG Count
0
25
50
75
100
Relative Abundance (%)
Phylum
Ascomycota – Primary Symbiont
Ascomycota – Additional Fungi
Chlorophyta
Cyanobacteriota
Acidobacteriota
Actinomycetota
Armatimonadota
Bacteroidota
Chloroflexota
Myxococcota
Planctomycetota
Pseudomonadota
Verrucomicrobiota
I
J
<1%
FACHB-36_1
Kovacikia_1
Stenomitos_1
Stenomitos_2
Nostoc_1
Nostoc_2
Nostoc_3
Nostoc_4
Nostoc_5
Nostoc_6
MAG Presence
MAG ID
FACHB-36_1
Kovacikia_1
Stenomitos_1
Stenomitos_2
Nostoc_1
Nostoc_2
Nostoc_3
Nostoc_4
Nostoc_5
Nostoc_6

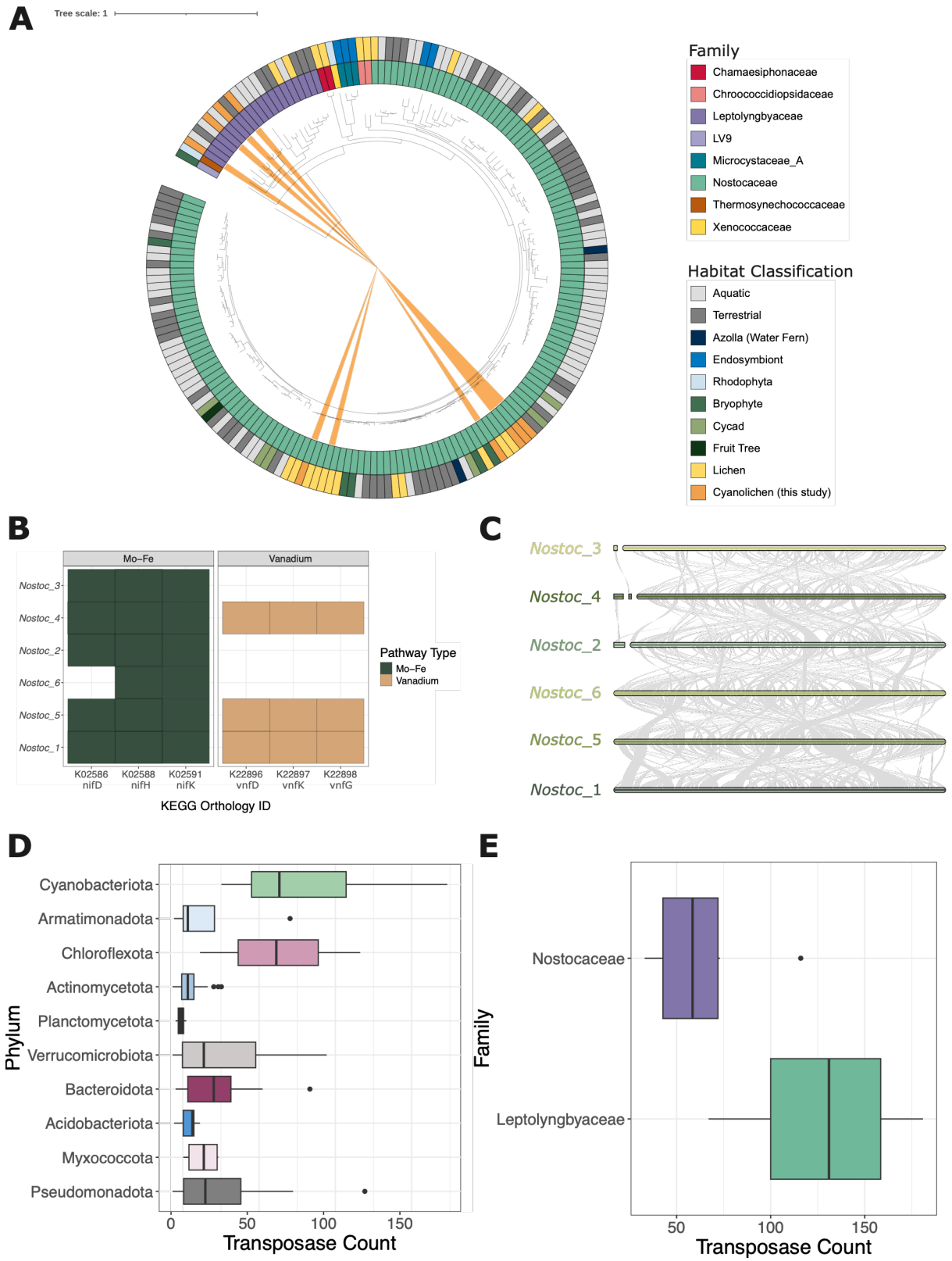
316

317 **Figure 1: Organismal composition of 11 different species of cyanolichens**
318 **determined using long-read metagenomic approaches.** (A-F) Thallus of lichen
319 species which include cyanobacterial photobionts sequenced in this study including
320 Leptogium burgessii (A), Lobaria pulmonaria (B), Ricasolia virens (C),
321 Pseudocyphellaria norvegica (D), Sticta sylvatica (E), and Peltigera collina (F). (G)
322 Absolute count of detected metagenomic assembled genomes (MAGs) and (H)
323 relative contribution of different taxa to cyanolichen metagenomic data generated.
324 Cyanolichen species shown in turquoise (Lobaria pulmonaria and Ricasolia virens)
325 are tripartite cyanolichens including both chlorophyte and cyanobacterial
326 photobionts. (I) Detection of cyanobacterial MAGs belonging to Nostocaceae (green)
327 and Leptolyngbyaceae (purple) across cyanolichen species. Cells marked with a '*'
328 indicate that nitrogen fixation pathways were detected in the genome. Genomes
329 accounting for <1% composition of the cyanobacterial component of lichen
330 metagenomes are indicated. (J) Relative composition of cyanobacterial components
331 in the sequenced lichen metagenomes.
332
333 Ten bacterial phyla were represented across the 103 bacterial MAGs:
334 Acidobacteriota, Actinomycetota, Armatimonadota, Bacteroidota, Chloroflexota,
335 Cyanobacteriota, Myxococcota, Planctomycetota, Pseudomonadota and
336 Verrucomicrobiota. Representatives from bacterial families commonly found in
337 lichens including Acidobacteriaceae, Beijerinckiaceae, Acetobacteraceae and
338 Sphingomonadaceae were detected (Figure S1). The diversity and frequency of
339 bacterial MAGs, excluding the cyanobacterial photobiont, between samples ranged
340 from 31 in Peltigera collina to just 3 in Peltigera membranacea. 77 MAGs were only
341 detected in a single sample (Figure S2; Table S17). Bacteria, excluding the
342 Cyanobacteriota, were cumulatively present in lower abundances in each lichen
343 metagenome in comparison to the primary mycobiont and photobiont (including both
344 chlorophyte and cyanobacteria) MAGs (Figure 1H).
345
346 Sequencing depths normalized as a function of genome/bin size were used to
347 explore the taxonomic composition of the cyanolichen metagenome. The lichenized
348 fungi (Ascomycota) were major components of the metagenome. Cyanobacteriota,
349 representative of the cyanobacterial photobiont in the lichen symbioses, were
350 frequently found to be dominant representatives within the lichen metagenome

accounting for up to 80.1% by copy number (see Methods) in Sticta sylvatica, and on average comprising 51.6% of metagenomic assemblies (Figure 1H). The relative abundance of cyanobacterial photobionts in the two tripartite cyanolichens, Lobaria pulmonaria and Ricasolia virens, was significantly less (0.77% and 2.5% respectively) and chlorophyte photobionts were more abundant (10.1% and 18.1%).

3.2 Alternative vanadium-dependent nitrogen-fixation pathways observed in Nostoc symbionts

Of the 103 bacterial MAGs identified across the 11 species of cyanolichens, 10 were identified as cyanobacteria. These 10 cyanobacterial MAGs included six genomes identified as distinct *Nostoc* species, and 4 genomes from the family Leptolyngbyaceae. *Nostoc* species are common cyanobacterial photobionts in lichens, but Leptolyngbyaceae have not previously been reported as lichen symbionts (Figure 1I). Further exploration of the Cyanobacteriota composition revealed that Leptolyngbyaceae MAGs were present in extremely low abundance when detected (maximum 3.3% in *Sticta sylvatica*; <1% in many other instances; Figure 1J) suggesting that these cyanobacteria are unlikely to be primary photobionts. The *Nostoc* strains identified in this study showed similarity to other previously identified *Nostoc* symbionts, while Leptolyngbyaceae taxa showed similarity to references originating in aquatic sources and not from symbiotic lineages (Figure 2A).


A
Tree scale: 1
Family
Chamaesiphonaceae
Chroococcidiopsidaceae
Leptolyngbyaceae
LV9
Microcystaceae_A
Nostocaceae
Thermosynechococcaceae
Xenococcaceae
Habitat Classification
Aquatic
Terrestrial
Azolla (Water Fern)
Endosymbiont
Rhodophyta
Bryophyte
Cycad
Fruit Tree
Lichen
Cyanolichen (this study)
B
Mo–Fe
Vanadium
Nostoc_3
Nostoc_4
Nostoc_2
Nostoc_6
Nostoc_5
Nostoc_1
K02586 nifD
K02588 nifH
K02591 nifK
K22896 vnfD
K22897 vnfK
K22898 vnfG
KEGG Orthology ID
Pathway Type
Mo–Fe
Vanadium
C
Nostoc_3
Nostoc_4
Nostoc_2
Nostoc_6
Nostoc_5
Nostoc_1
D
Phylum
Cyanobacteriota
Armatimonadota
Chloroflexota
Actinomycetota
Planctomycetota
Verrucomicrobiota
Bacteroidota
Acidobacteriota
Myxococcota
Pseudomonadota
0
50
100
150
Transposase Count
E
Family
Nostocaceae
Leptolyngbyaceae
50
100
150
Transposase Count

373

374 **Figure 2: Diversity, distribution and genomic exploration of cyanobacterial**
375 **metagenomic assembled genomes (MAGs).** (A) Phylogenomic tree of
376 dereplicated Nostocaceae and Leptolyngbyaceae genomes to identify conservation
377 of symbiont types. Orange shaded branches of the tree indicate cyanobacterial
378 MAGs generated in this study. (B) Detection of KEGG orthologs in molybdenum-iron
379 (Mo-Fe) and vanadium dependent nitrogen fixation pathways in *Nostoc* MAGs. *vnfH*
380 was not detected in any *Nostoc* MAG and is not shown. (C) Synteny plots of *Nostoc*
381 genomes to identify areas of genomic rearrangements. (D) Frequency count of
382 transposase genes in bacterial MAGs. (E) Frequency count of transposase genes in
383 cyanobacterial MAGs only.
384
385 We found at least one nitrogen-fixing cyanobacterial symbiont per lichen thallus.
386 However, in some cases, such as *Ricasolia virens*, multiple different species of
387 nitrogen-fixing *Nostoc* were identified from a single thallus sample (Figure 1I). While
388 some cyanolichens were found to host multiple species of *Nostoc*, one species was
389 always found to be dominant in all cyanolichens assessed. Nitrogen-fixation
390 pathways in these *Nostoc* symbionts included the molybdenum-iron dependent
391 pathway in all *Nostoc* MAGs. Genes required for the vanadium dependent pathway
392 were detected in three out of six *Nostoc* species, but the pathways were potentially
393 incomplete as the vnfH gene was not detected (Figure 2B). Syntenic analysis of
394 orthologous regions of *Nostoc* symbiont MAGs showed large rearrangements
395 between different genomes (Figure 2C). Cyanobacterial MAGs were shown to have
396 significantly more transposases than other bacteria detected in the cyanolichen
397 (Figure 2D; Table S18). More transposases were identified in Leptolyngbyaceae
398 than Nostocaceae MAGs (Figure 2E).
399
400 *3.3 Transposable elements drive genome expansion in chlorophyte symbiont*

401 The chlorophyte symbiont from the tripartite cyanolichens, *Ricasolia virens* and
402 *Lobaria pulmonaria*, was recovered independently in each sample but was shown to
403 be the same species based on average nucleotide identity (ANI > 0.99; Table S19).
404 Phylogenetic placement using the 18S rRNA gene showed relatedness to the
405 trebouxioid symbiont, *Symbiochloris*. This identification was further confirmed with
406 BLAST results (Figure 3A). The nuclear genome size was 75.8 Mb in 31 contigs

(Figure 3B), with a contig N50 of 3.5 Mb, and estimated BUSCO completeness of 90.8%. Circularized organellar genomes for chloroplast and mitochondrion were recovered with sizes of 354,908bp and 105,833bp respectively (Figure S3/4).

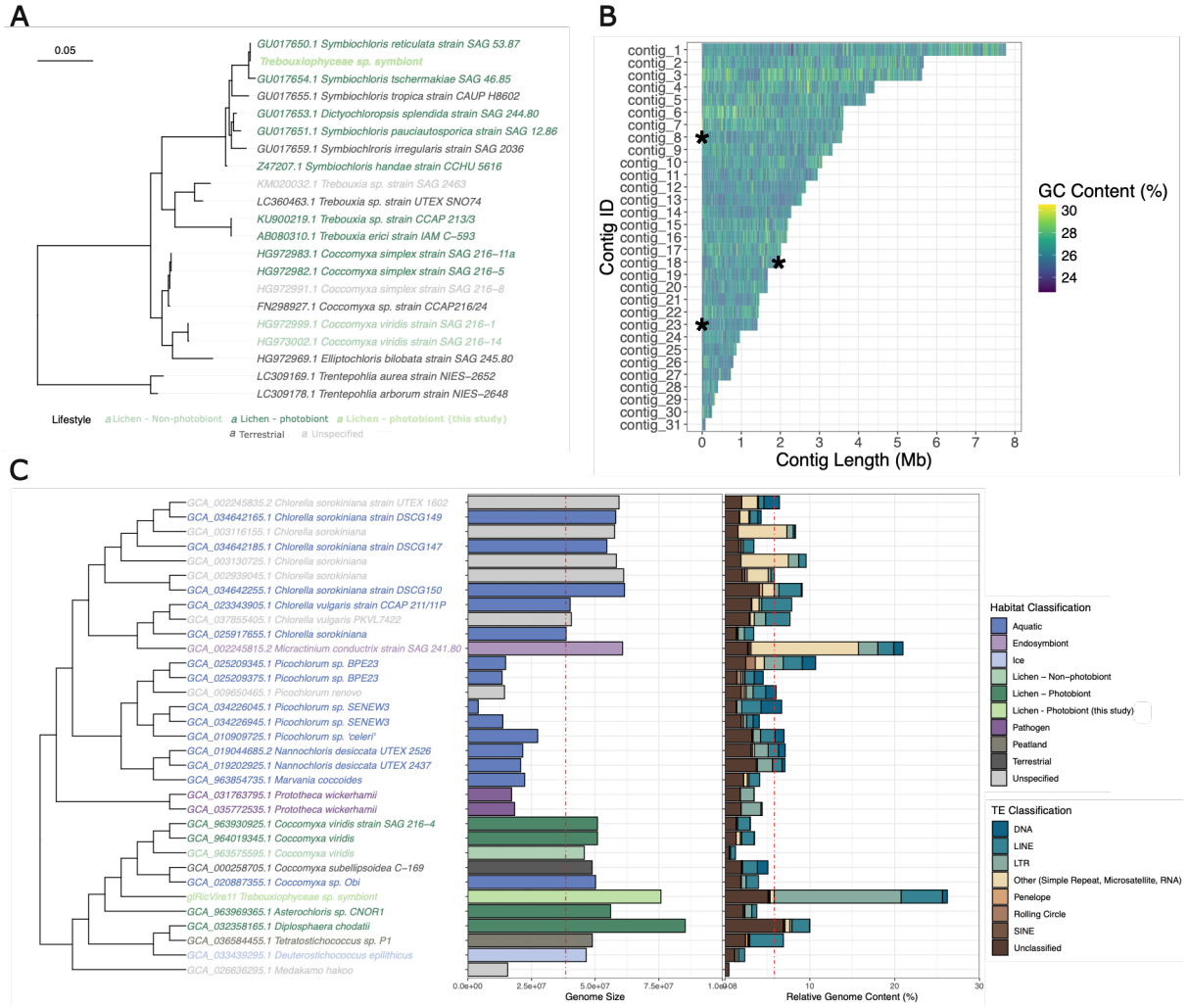


Figure 3: Identification and genomic characteristics of the chlorophyte photobiont in *Ricasolia virens* and *Lobaria pulmonaria*. (A) Phylogenetic tree based on the 18S rRNA locus showing taxonomic placement of recovered chlorophyte photobionts from the tripartite cyanolichens *Lobaria pulmonaria* and *Ricasolia virens* in the context of other free-living chlorophyte algae and chlorophyte symbionts. (B) GC content across contig length of the Trebouxiophyceae symbiont genome. Contig ends marked with a * indicate the detection of telomeres at the start (left-handed placement) or end (right-handed placement) of the contig. (C) Genome size and transposable element content of chlorophyte photobiont and other Trebouxiophyceae from different habitat types. Median genome size and transposable element content indicated with a dashed red line.

423 The genome size of the chlorophyte symbiont recovered in this study (75.8Mb) was
424 larger than most other Trebouxiophyceae genomes (median=46.1Mb) used in our
425 comparative analyses. Characterization of the transposable elements of the
426 chlorophyte symbiont MAG revealed the genome to be comprised of 26% TE content
427 (Figure 3B) and identifies this as one of the most repeat-rich Trebouxiophyceae
428 genomes in comparison to a median TE content of 3.6% (Figure 3C). Kimura
429 distances revealed the repeat landscape in this chlorophyte symbiont to be acquired
430 recently in the evolutionary history of this genome, with similar Kimura distances
431 observed in other trebouxioid symbionts and free-living strains (Figure S5).
432
433 Genes were assigned to categorical groups based on their distance to TEs ranging
434 from overlapping (0 bp) to distant (>20,000 bp). We found 4156 genes overlapping
435 with TE annotations in the Trebouxiophyceae symbiont from *Ricasolia virens* and
436 *Lobaria pulmonaria* but only 65% (n = 2700) were detectably expressed (Figure 4A).
437 Despite the same species of Trebouxiophyceae symbiont being identified in both
438 *Ricasolia virens* and *Lobaria pulmonaria* differences in gene expression in proximity
439 to TEs were observed between these different lichens. A lower proportion of genes,
440 including those overlapping or in close proximity to TEs were, were observed to be
441 detectably expressed in the Trebouxiophyceae symbiont of *Lobaria pulmonaria*
442 (59.5%) than the symbiont of *Ricasolia virens* (77.5%) suggesting dynamic
443 regulation of genes between individual lichen samples.
444
445 GO term enrichment for TE distance categories revealed variable enriched functions
446 in Trebouxiophyceae symbionts independent of proximity to TEs and between
447 lichens (Figure 4B; Table S20). Terms associated with essential cellular processes
448 were found enriched at further distance from TEs with the exception of
449 photosynthesis (Figure 4C). Stress responses and functions which may be
450 associated with environmental communication were found to be enriched in the
451 categorical groups corresponding to closer proximity TEs (1-1000bp; 1001-5,000bp;
452 Figure 4D). In addition to these enriched terms, specific functions associated with
453 chlorophyte lichenization were frequently found in close proximity to TEs (<10,000bp;
454 Figure 4E).
455

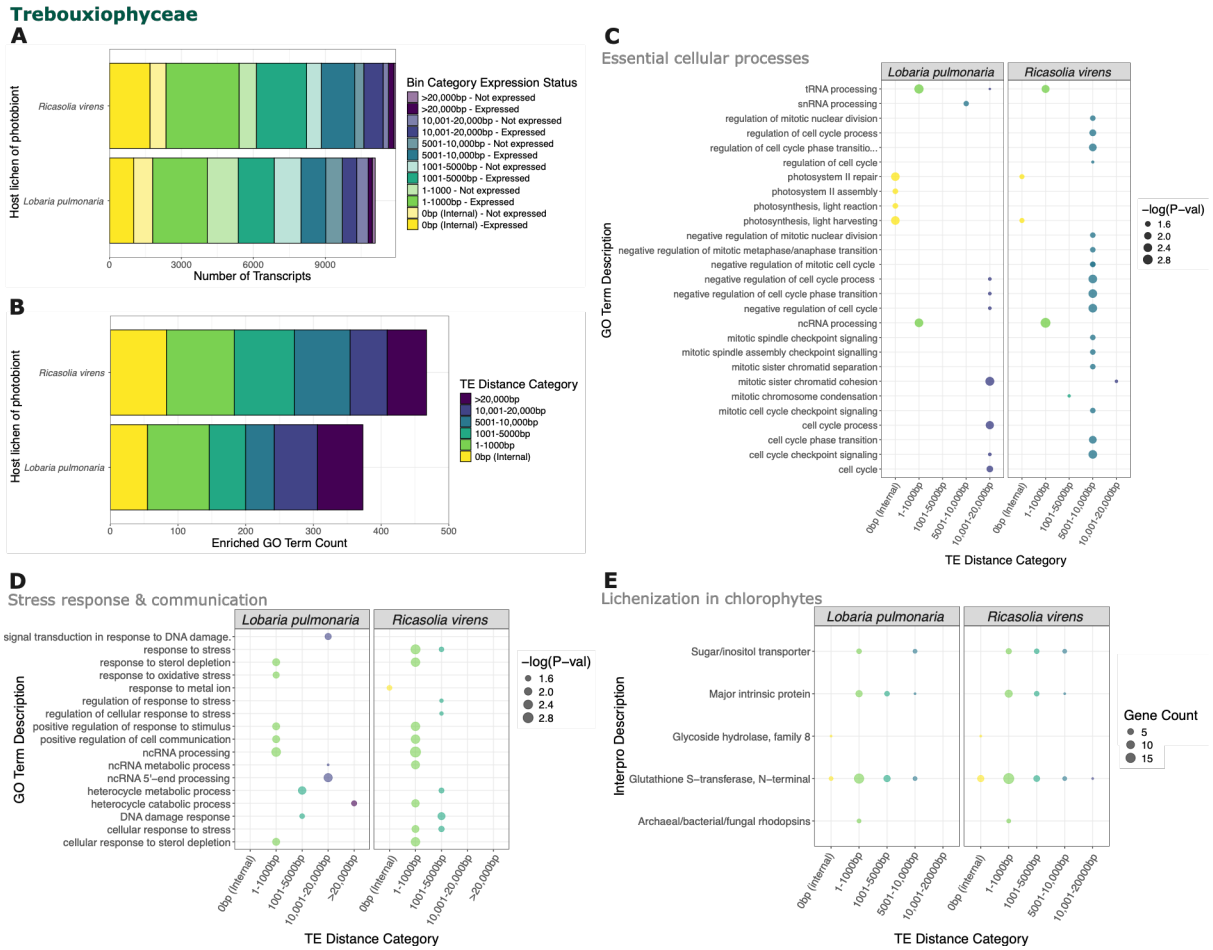


Figure 4: Trebouxiophyceae (*Symbiochloris*) symbiont gene expression and functional enrichment in tripartite cyanolichens, *Lobaria pulmonaria and Ricasolia virens*. (A) Expression status of genes in proximity to transposable elements shows that many genes are in close proximity to transposable elements with detectable expression in Trebouxiophyceae symbiont MAGs. (B) Frequency of enriched GO terms in Trebouxiophyceae symbiont MAGs in relation to nearest distance to transposable element. (C-D) Enrichment of GO terms in proximity to nearest TE, where size of dot represents -log(p-value). (C) Functions associated with essential cellular processes such as the cell cycle and RNA processing are observed to be enriched in genes at further distances from TE with exception of photosynthesis related GO terms. (D) Enriched GO terms show enrichment of functions corresponding to stress response and environmental communication in genes closer to transposable elements. (E) Proximity of selected genes previously associated with lichenization in chlorophytes to TEs where size of dot represents gene counts.

3.4 Metagenomic recovery of largest Lecanormycetes genomes to date

Of the 15 fungal MAGs, 11 correspond to the primary fungal symbiont of the lichenized fungi from the taxonomic order Peltigerales. An additional four ascomycete fungi were identified in the cyanolichens. MAGs for the primary fungal symbionts of the Peltigerales fungi showed genomes ranging from 30.98 Mb (*Leptogium burgsessii*) to 137 Mb (*Peltigera hymenina*) (Table S9). *Peltigera* species sequenced in this study have the largest reported genome sizes for Lecanormycetes fungi characterized to date, which previously had a recorded median genome size of 66.4 Mb (Figure 5A).

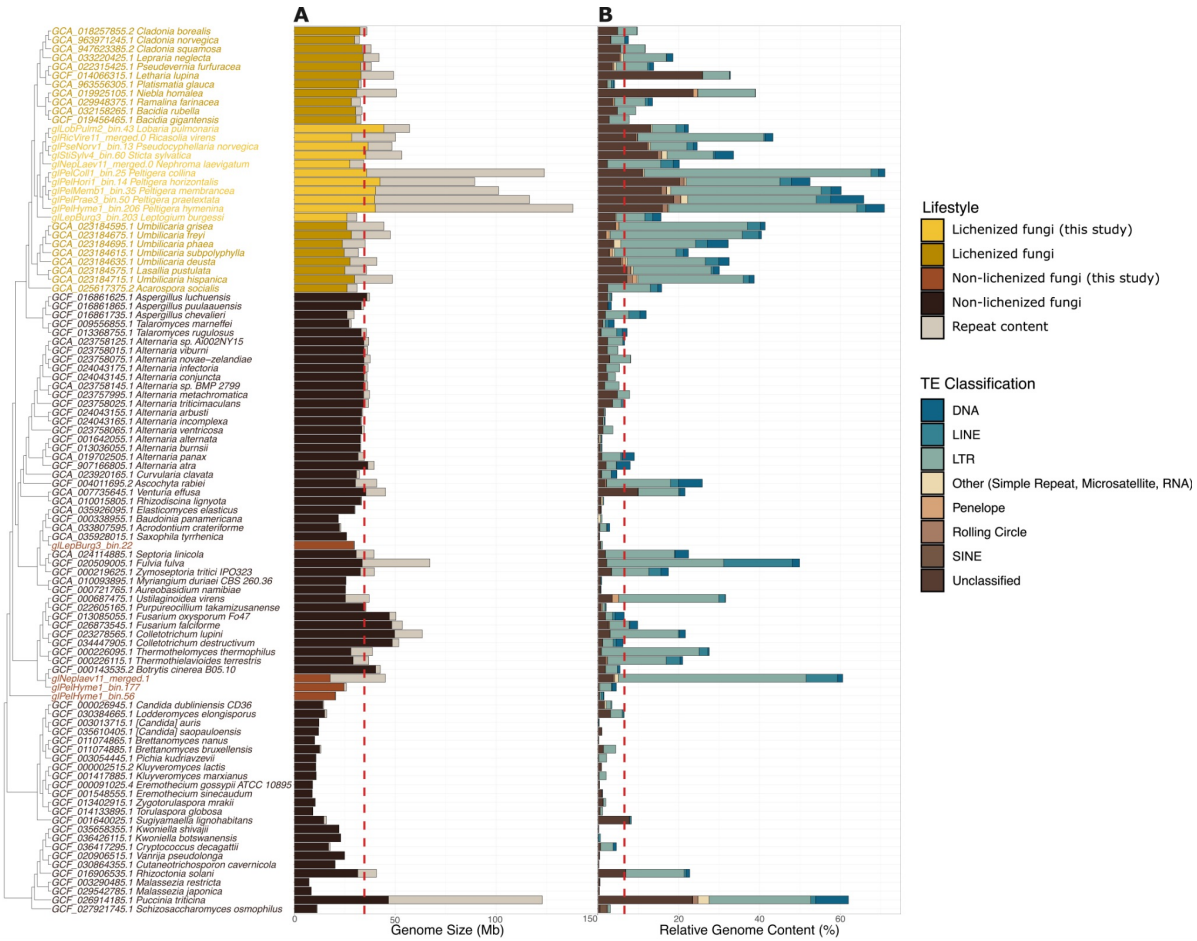


Figure 5: Genome size and repeat content of lichenized and non-lichenized fungi. (A) Genome size and (B) transposable element content of fungi arising from both lichenized and non-lichenized sources, including novel MAGs generated in this study. Median values of (A) genome size and (B) transposable element content (%) indicated with red dashed line.

The elevated genome sizes observed in the Peltigerales in this study may be attributed to the high TE content, which accounts for up to 71% of the genome span. These Pelitgerales are the most repeat-rich lichenized fungi genomes, and have a significantly higher repeat content than other non-lichenized fungi (Figure 5B). Contrasting the Kimura distance profiles of the TEs of the lichenized and non-lichenized fungal MAGs generated in this study showed stark differences. Peltigerales fungi have had consistent and steady TE acquisition throughout their evolutionary history, while non-lichenized fungi have comparatively shorter and smaller acquisition events (Figure S6). Transposable elements of non-lichenized fungi were generally located in broad heterochromatin-like regions, as opposed to TEs in lichenized fungi which were more commonly found in smaller TE-rich regions interspersed throughout the contigs (adjusting for genome size; Figure 6A). Gene expression levels were marginally affected by their presence within gene-rich, TE-rich, or mixed regions of the genome (Figure 6B). As a result of this pattern of TEs in the lichenized fungi genomes, many genes were found to be in close proximity to transposable elements (Figure 7A). Potential TE fusion candidates were identified based on protein domain composition (Figure S7/S8) which were shown to be actively expressed in the Peltigerales lichenized fungi (Figure S9).

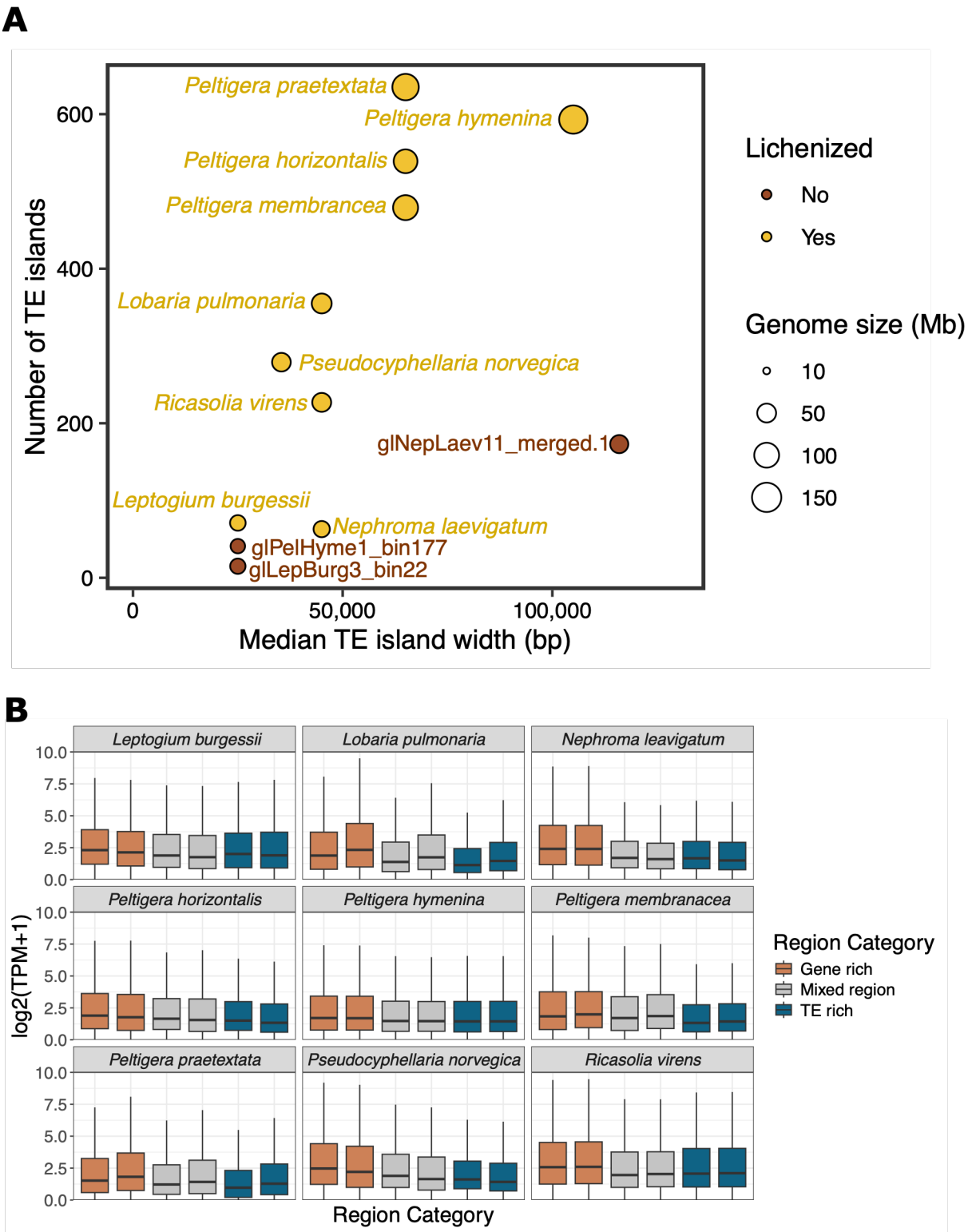


Figure 6: Distribution and expression of genes and TEs in fungal genomes. (A) Scatter plot showing the relationship between the number of TE-enriched regions (TE islands) and their size (TE island width), comparing with the genome size (point size) of the fungal genomes. (B) Boxplots of log TPM expression values of genes present within gene-rich, TE-rich, or mixed regions in the genomes of all sequenced

515 lichenized fungi shown between two replicates. The average expression of genes
516 within gene-rich regions is only slightly higher than the average expression of genes
517 found within TE-rich regions.
518
519 *3.5 Cellular stress response genes enriched near transposable elements*

520 Following annotation, 23,360 genes across 9 species of cyanolichen (Leptogium
521 burgessii, Lobaria pulmonaria, Peltigera praetextata, Peltigera horizontalis, Peltigera
522 hymenina, Peltigera membranacea, Ricasolia virens, Nephroma laevigatum,
523 Pseudocyphellaria norvegica) were identified to be overlapping with transposable
524 elements with 76% (n = 17,770) detectably expressed in RNA-seq data suggesting
525 ongoing TE activity (Figure 7B). The closest TEs to genes that were actively
526 expressed included DNA transposons (Fot1 and Sagan of Mariner class), long-
527 terminal repeat (LTR) retrotransposons (Gyspy and Copia), long interspersed
528 nuclear elements (LINE; L1 and Tad1), satellites and unclassified elements (Figure
529 S10). DNA transposons were present significantly closer to genes on average
530 compared to other TE families in all species, while LTR retrotransposons were
531 significantly further away in all species (Figure S11). While a high number of genes
532 were found to be actively expressed even when present in TE-rich islands of the
533 genome, the overall expression of these genes was comparable, if slightly lower on
534 average compared to genes within gene-rich regions (Figure 6B; Figure S12).
535
536 Genes were assigned to categorical groups based on their distance to TEs ranging
537 from overlapping (0 bp) to distant (>20,000 bp). GO term enrichment for TE distance
538 categories revealed enriched functions in fungal genomes dependent on proximity to
539 TE (Figure 7C; Table S21). Genes, globally, were further away from TEs than
540 expected by chance (Table S22). Despite this, genes overlapping or in close
541 proximity (<1000 bp) to TEs were found to be enriched for cellular responses to
542 stress and stimulus (e.g., radiation, UV, abiotic stress), transmembrane transport,
543 and biosynthesis and metabolism of secondary metabolites including mycotoxins
544 and heterocycles (Figure 8A). Genes at a greater distance from TEs were enriched
545 for GO terms associated with essential cellular processes such as RNA processing
546 and cell cycle maintenance (Figure 8B), suggesting stronger selection against TE
547 invasion of such genes. Similar results were obtained using permutation tests,

shuffling the locations of the genes to determine whether any GO terms were associated with genes closer or further away from TEs than expected (Figure S13). Semantic enrichment and clustering revealed some distinct terms enriched between those significantly closer or further away from TEs (Figure S14/S15).

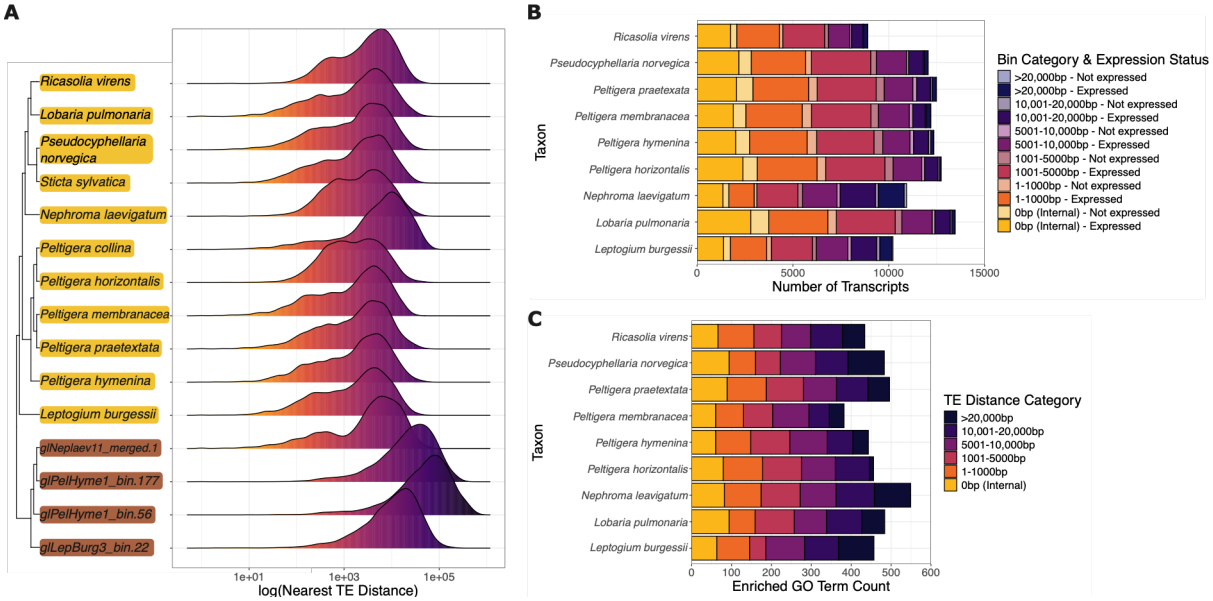


Figure 7: Expression and functional enrichment of genes in relation to nearest transposable element in fungal metagenomic assembled genomes (MAGs). (A) Density plot of the distance to nearest transposable element for lichenized and non-lichenized MAGs. Lichenized fungi taxa are highlighted in gold and non-lichenized in brown. (B) Expression status of genes in proximity to transposable elements shows that many genes are in close proximity to transposable elements while being actively expressed in lichenized fungal MAGs. (C) Frequency of enriched GO terms in lichenized fungi MAGs in relation to nearest distance to transposable element.

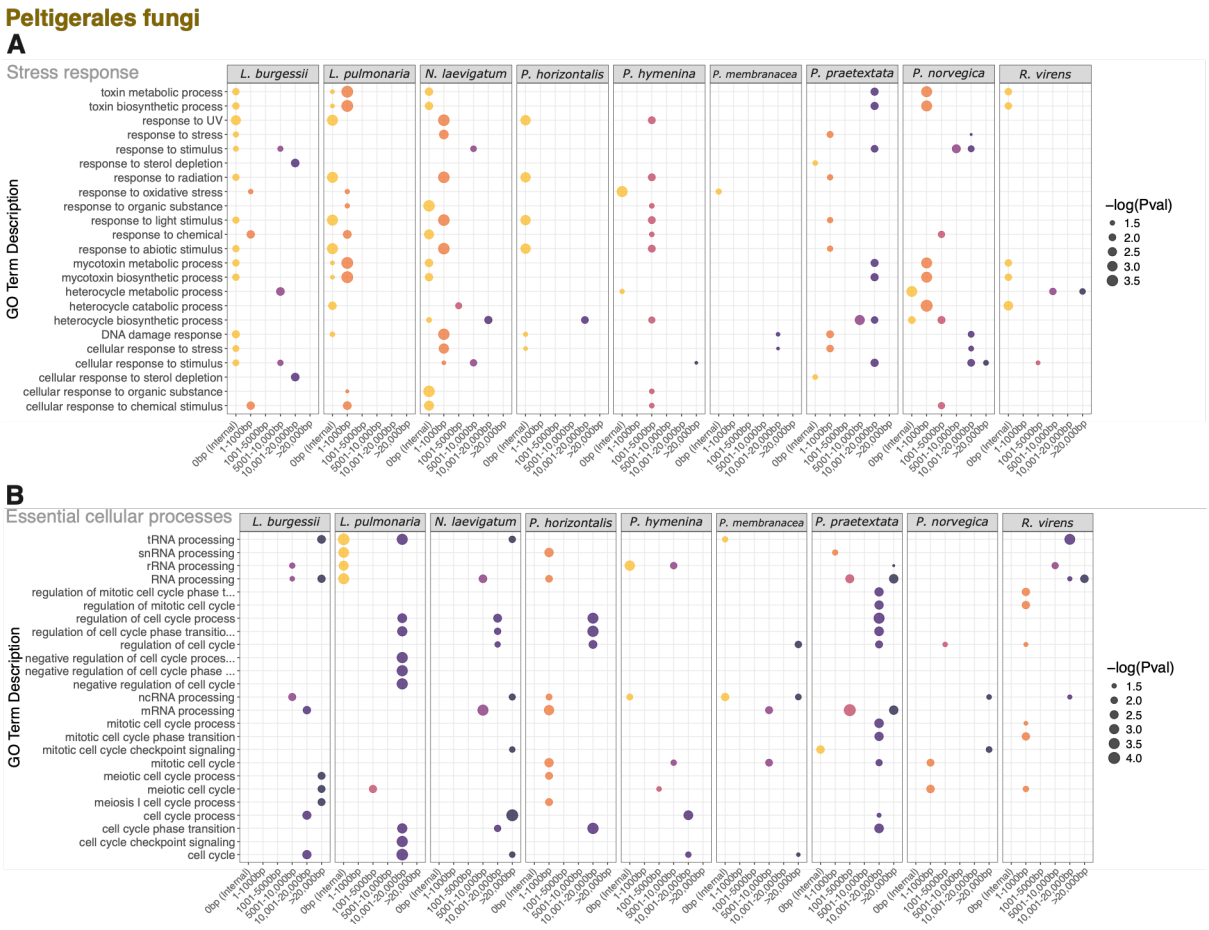


Figure 8: Enriched GO terms in Peltigerales fungal metagenomic assembled genomes (MAGs). Enriched GO terms corresponding to stress response and environmental communication (A) and essential cellular processes (B) where the point size represents -log(p-value). GO terms corresponding to stress response were observed in genes closer to TEs while those associated with essential cellular processes were at further distances from TEs.

4. Discussion

We performed long-read metagenomic sequencing of 11 species of cyanolichens to explore organismal composition and genomic content in the symbiotic association. The lichen-associated bacterial communities were found to contain key lineages that have previously been reported in lichens[9] (Figure 1G). We found frequent co-occurrence of multiple species of *Nostoc* within the same lichen (Figure 1I) emphasising previous demonstrations of cyanobacterial photobiont distribution and promiscuity[7,66,67]. Some *Nostoc* symbionts were also likely to possess an alternative vanadium dependent nitrogen fixation pathway (Figure 2B) as previously observed in

579 other cyanobacterial symbionts[7,68–70] suggesting a broader specialization to support
580 nitrogen fixation independent of environmental conditions. Genomic rearrangements
581 and high counts of transposase genes in the cyanobacterial genomes (Figure 2D)
582 also suggest the potential impact of mobile genetic elements in essential bacterial
583 symbionts to allow for host adaptation[71].
584
585 The ability to explore the genomics of the eukaryotic symbionts in lichens has
586 frequently lagged behind the exploration of their bacterial counterparts due to the
587 challenge of recovering high-quality, highly contiguous MAGs from short-read
588 metagenomic datasets. Despite the bulk of biomass arising from the fungal partner,
589 previous metagenomic studies using short-read sequencing have faced challenges
590 in recovery of both fungal and algal MAGs[9] . This has prevented advances in the
591 understanding of lichen evolution as the number of publicly available lichenized
592 fungal genomes differs by an order of magnitude from current estimated species
593 counts. In this study, we generated highly contiguous, high quality fungal genomes
594 for 15 fungi, 11 of which are lichenized, presenting the opportunity to explore
595 genomic characteristics and specialized molecular functions in the lichenized
596 lineage, Peltigerales.
597
598 The dependence of lichenized fungi on their photobiont for growth draws parallels to
599 the lifestyle features observed in obligate biotrophic pathogenic fungi, which can only
600 grow on living host tissue. We observe similar trends in genome size and TE content
601 between these two groups of fungi with co-dependent lifestyles[20]. Many of the largest
602 fungal genomes characterized to date arise from obligate biotrophic pathogens with
603 the largest, *Austropuccinia psidii* with a genome of 1 Gb, being composed of >90%
604 TEs[19] . While not as extreme, the Peltigerales lichens sequenced in this study have
605 above-average genome sizes that can be attributed to high TE content and
606 represent the most TE-rich genomes of lichenized fungi characterised to date (Figure
607 5). Transposable element content in the context of fungal lifestyles has been
608 explored previously[72,73] and the trends observed here add a further layer to the
609 impact of lifestyle on TE content: the degree of dependency of fungi on other
610 organisms (e.g. obligate lifestyles) appears to be associated with larger genome
611 sizes driven by increased TE content.

612 The biological impact of increased numbers of TEs and their distribution in a genome
613 is likely complex. TEs can influence the expression of genes near insertion sites, by
614 adding potential promoters and enhancers, and by impacting expression in local
615 domains when they are epigenetically silenced. We explored whether genes with
616 annotations overlapping or near TE annotations were more or less likely to be
617 expressed. We confirm that genes overlapping with and nearby to TEs in lichenized
618 fungi are actively expressed at levels comparable to other genes and that these
619 genes can be associated with functions with biological relevance to the lichenized
620 lifestyle (Figure 8A). These results also evoke a similar phenomenon observed in
621 fungal pathogen TE evolution, where the copy number and diversity of actively
622 expressed virulence genes (often present near TEs) is positively associated with TE
623 expansion occurring in smaller TE-rich regions scattered throughout the genome[20].
624 This parallel could suggest that the ability of Peltigerales lichens to participate in the
625 symbiosis may be linked to such TE proximal genes undergoing rapid evolution.
626 When looking into genes in close proximity to TEs in the Peltigerales lichens, we
627 observed a high abundance of genes associated with responses to stress,
628 environmental stimuli and production of secondary metabolites which correspond to
629 many of the characteristics associated with lichenization and their ability to survive
630 harsh environments and produce diverse secondary metabolites. The samples used
631 in this study were desiccated and the transcriptomes reflected this metabolic status.
632 Future transcriptomic studies exploring the life cycle and desiccation processes in
633 lichens will be required to further elucidate the impact of TEs on gene regulation.
634
635 In the gradient of necessity of the symbiotic relationship in lichens, the lichenized
636 fungi within the Lecanoromycetes lean towards being obligate symbionts, leaving the
637 question remaining as to whether similar patterns of TE content and genome
638 expansion are observed in non-obligate symbionts, such as lichenized chlorophytes
639 or other fungal lineages. The TE content of algae has been underexplored with
640 estimates of TEs contributing to a small portion of the overall genome size in the
641 Chlorophyta. In our analysis of the Trebouxiophyceae symbiont, detected in both
642 *Ricasolia virens* and *Lobaria pulmonaria*, we show that some lichenized trebouxioid
643 photobionts may have elevated TE content, but this is often comparable to that
644 observed in free-living aquatic species. Recent research on lichenization in
645 chlorophytes revealed specialized carbohydrate active enzymes amongst other

646 functions associated with adaptation to high light intensity and drought tolerance[8,9].
647 In our analyses, genes associated with these functions were frequently found in
648 close proximity to TEs (Figure 4E; Figure S15). A carbohydrate active enzyme from
649 glycoside hydrolase family 8 that has been associated with the lichenization of
650 chlorophytes[8] was noted to contain a TE fragment within it. While this may play a
651 role in the functional adaptation of the algae, the absence of significant genome
652 expansions as observed in the lichenized fungi presented in this work could indicate
653 one of two scenarios: either the Trebouxiophyceae symbiont is under strong
654 pressure to control TE content and does not accept TE insertions, or the
655 Trebouxiophyceae symbiont is under less pressure to adapt and fewer genomic
656 changes, including TE expansions, are occurring.
657
658 The relationship of genes and TEs in lichen symbionts can broadly be classified into
659 five simplified scenarios: (1) genes are arising from domesticated TEs; (2) TEs are
660 impacting the expression of functional genes relevant to symbiosis and the
661 lichenized lifestyle; (3) novel functional genes essential to the lichenized lifestyle are
662 formed through the fusion of TE domains to genome domains; (4) genes are not
663 impacted by TE activity; or (5) the distribution of smaller TE-rich regions in the
664 genome is leading to an increase in copy number variation of symbiosis-related
665 genes. The work presented here constitutes the first exploration of TE expression
666 and activity in the Peltigerales fungi and their associated chlorophyte photobionts
667 and identifies a potential association of obligate versus facultative lifestyle on high
668 versus low TE content and TE distribution. As more metagenomic research utilises
669 long-read sequencing, the expected recovery of high-quality, highly-contiguous
670 eukaryotic MAGs will increase creating further opportunity to characterize TE content
671 in diverse microbial eukaryotes and population level genome dynamics to further
672 advance our understanding of these genomic elements on evolutionary processes.

673 **Acknowledgements**

674 We acknowledge the support of NatureScot for site access and sampling
675 permissions for field samples collected in Glen Creran Woods. This work was
676 supported by EMBL (European Molecular Biology Laboratory) core funds and
677 Wellcome Sanger Institute (220540/Z/20/A). This research was funded in part, by the

Wellcome Trust. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Data and code availability

The raw sequencing data generated in this project at the Wellcome Sanger Institute is publicly available under project accessions PRJEB63593 and PRJEB73802. Reference sequences used for phylogenetic placement and comparison available in public repositories with accession information in Supplementary Tables S12, S14, S15, S16. Custom scripts for data analysis and visualization are available at https://github.com/Finn-Lab/2026_transposable_elements_cyanolichens. Additional files required to reanalyze the data including annotation files and associated data frames available upon request.

5. References

1. Spribille, T., Resl, P., Stanton, D. E. & Tagirdzhanova, G. Evolutionary biology of lichen symbioses. *New Phytologist* **234**, 1566–1582 (2022).

2. Lücking, R., Hodkinson, B. P. & Leavitt, S. D. The 2016 classification of lichenized fungi in the Ascomycota and Basidiomycota – Approaching one thousand genera. *The Bryologist* **119**, 361–416 (2017).

3. Rikkinen, J., Oksanen, I. & Lohtander, K. Lichen guilds share related cyanobacterial symbionts. *Science* vol. 297 357 (2002).

4. Stenroos, S., Högnabba, F., Myllys, L., Hyvönen, J. & Thell, A. High selectivity in symbiotic associations of lichenized ascomycetes and cyanobacteria. *Cladistics* **22**, 230–238 (2006).

5. Hawksworth, D. L. Lichenization: The Origins of a Fungal Life-Style. in *Recent Advances in Lichenology: Modern Methods and Approaches in Lichen Systematics and Culture Techniques, Volume 2* (eds Upreti, D. K., Divakar, P. K., Shukla, V. & Bajpai, R.) 1–10 (Springer India, New Delhi, 2015). doi:10.1007/978-81-322-2235-4_1.

6. Gargas, A., DePriest, P. T., Grube, M. & Tehler, A. Multiple Origins of Lichen Symbioses in Fungi Suggested by SSU rDNA Phylogeny. *Science* **268**, 1492–1495 (1995).

7. Cameron, E. S., Sanchez, S., Goldman, N., Blaxter, M. L. & Finn, R. D. Diversity and specificity of molecular functions in cyanobacterial symbionts. *Sci Rep* **14**, 18658 (2024).

8. Puginier, C. *et al.* Phylogenomics reveals the evolutionary origins of lichenization in chlorophyte algae. *Nat Commun* **15**, 4452 (2024).

9. Tagirdzhanova, G. *et al.* Microbial occurrence and symbiont detection in a global sample of lichen metagenomes. *PLOS Biology* **22**, e3002862 (2024).

10. Grimm, M. *et al.* The Lichens' Microbiota, Still a Mystery? *Frontiers in Microbiology* vol. 12 (2021).

11. Smith, H. B. *et al.* Metagenomic data reveal diverse fungal and algal communities associated with the lichen symbiosis. *Symbiosis* **82**, 133–147 (2020).

12. Spribille, T. *et al.* Basidiomycete yeasts in the cortex of ascomycete macrolichens. *Science* **353**, 488–492 (2016).

13. Merges, D., Dal Grande, F., Greve, C., Otte, J. & Schmitt, I. Virus diversity in metagenomes of a lichen symbiosis (Umbilicaria phaea): complete viral genomes, putative hosts and elevational distributions. *Environmental Microbiology* **23**, 6637–6650 (2021).

14. Scharnagl, K., Tagirdzhanova, G. & Talbot, N. J. The coming golden age for lichen biology. *Current Biology* **33**, R512–R518 (2023).

15. Kojima, K. K. Structural and sequence diversity of eukaryotic transposable elements. *Genes Genet Syst* **94**, 233–252 (2020).

16. Wells, J. N. & Feschotte, C. A Field Guide to Eukaryotic Transposable Elements. *Annu Rev Genet* **54**, 539–561 (2020).

17. Castanera, R. *et al.* Transposable Elements versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. *PLOS Genetics* **12**, e1006108 (2016).

18. Dallaire, A. *et al.* Transcriptional activity and epigenetic regulation of transposable elements in the symbiotic fungus Rhizophagus irregularis. *Genome Res* **31**, 2290–2302 (2021).

19. Tobias, P. A. *et al.* Austropuccinia psidii, causing myrtle rust, has a gigabase-sized genome shaped by transposable elements. *G3 Genes|Genomes|Genetics* **11**, jkaa015 (2021).

741 20. Torres, D. E., Oggenfuss, U., Croll, D. & Seidl, M. F. Genome evolution in
742 fungal plant pathogens: looking beyond the two-speed genome model. *Fungal*
743 *Biology Reviews* **34**, 136–143 (2020).
744 21. Liu, P., Cuerda-Gil, D., Shahid, S. & Slotkin, R. K. The Epigenetic Control of
745 the Transposable Element Life Cycle in Plant Genomes and Beyond. *Annu Rev*
746 *Genet* **56**, 63–87 (2022).
747 22. Seidl, M. F. & Thomma, B. P. H. J. Transposable Elements Direct The
748 Coevolution between Plants and Microbes. *Trends Genet* **33**, 842–851 (2017).
749 23. Dal Grande, F. *et al.* Transposable Elements in the Genome of the Lichen-
750 Forming Fungus Umbilicaria pustulata and Their Distribution in Different Climate
751 Zones along Elevation. *Biology* **11**, 24 (2022).
752 24. Tagirdzhanova, G. *et al.* Tangerine: a Starship-like element in the genomes of
753 Xanthoria lichen-forming fungi. 2025.11.25.690456 Preprint at
754 https://doi.org/10.1101/2025.11.25.690456 (2025).
755 25. Park, S.-Y., Jang, S.-H., Oh, S.-O., Kim, J. A. & Hur, J.-S. An Easy, Rapid,
756 and Cost-Effective Method for DNA Extraction from Various Lichen Taxa and
757 Specimens Suitable for Analysis of Fungal and Algal Strains. *Mycobiology* **42**, 311–
758 316 (2014).
759 26. Benoit, G. *et al.* High-quality metagenome assembly from long accurate reads
760 with metaMDBG. *Nature Biotechnology* https://doi.org/10.1038/s41587-023-01983-6
761 (2024) doi:10.1038/s41587-023-01983-6.
762 27. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*
763 **34**, 3094–3100 (2018).
764 28. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and
765 composition. *Nature Methods* **11**, 1144–1146 (2014).
766 29. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and
767 efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359
768 (2019).
769 30. Saary, P., Mitchell, A. L. & Finn, R. D. Estimating the quality of eukaryotic
770 genomes recovered from metagenomic analysis with EukCC. *Genome Biology* **21**,
771 (2020).
772 31. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov,
773 E. M. BUSCO: assessing genome assembly and annotation completeness with
774 single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

775 32. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for
776 genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
777 33. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and
778 accurate genomic comparisons that enables improved genome recovery from
779 metagenomes through de-replication. *ISME J* **11**, 2864–2868 (2017).
780 34. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: A
781 toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* **36**,
782 1925–1927 (2020).
783 35. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W.
784 CheckM: Assessing the quality of microbial genomes recovered from isolates, single
785 cells, and metagenomes. *Genome Research* **25**, 1043–1055 (2015).
786 36. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*
787 **30**, 2068–2069 (2014).
788 37. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification.
789 *Bioinformatics* **30**, 1236–1240 (2014).
790 38. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin,
791 L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat*
792 *Methods* **14**, 587–589 (2017).
793 39. Nguyen, L. T., Schmidt, H. A., Haeseler, A. V. & Minh, B. Q. IQ-TREE: A fast
794 and effective stochastic algorithm for estimating maximum-likelihood phylogenies.
795 *Molecular Biology and Evolution* **32**, 268–274 (2015).
796 40. Xu, S. *et al.* Ggtree: A serialized data object for visualization of a phylogenetic
797 tree and annotation data. *iMeta* **1**, e56 (2022).
798 41. Tang, H. *et al.* JCVI: A versatile toolkit for comparative genomics analysis.
799 *iMeta* **3**, e211 (2024).
800 42. Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile
801 HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
802 43. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the
803 phylogenetic tree display and annotation tool. *Nucleic Acids Res* **52**, W78–W82
804 (2024).
805 44. Tagirdzhanova, G. *et al.* Multipartite complexity of the lichen symbiosis
806 revealed by metagenome and transcriptome analysis of Xanthoria parietina.
807 2024.08.16.608140 Preprint at https://doi.org/10.1101/2024.08.16.608140 (2024).

808 45. Tagirdzhanova, G., Raistrick, J. & Talbot, N. J. Chromosome-level genome
809 assembly of the photobiont microalga Trebouxia sp. 'A48' from the lichen Xanthoria
810 parietina. *New Phytologist* **n/a**,.
811 46. Tremblay, B. J.-M. universalmotif: An R package for biological motif analysis.
812 *Journal of Open Source Software* **9**, 7012 (2024).
813 47. Seemann, T. barrnap 0.9: rapid ribosomal RNA prediction. *Google Scholar*
814 **792**, (2013).
815 48. Karlicki, M., Antonowicz, S. & Karnkowska, A. Tiara: deep learning-based
816 classification system for eukaryotic sequences. *Bioinformatics* **38**, 344–350 (2022).
817 49. Cameron, E. S., Blaxter, M. L. & Finn, R. D. plastiC: A pipeline for recovery
818 and characterization of plastid genomes from metagenomic datasets. *Wellcome*
819 *Open Research* **8**, 475 (2023).
820 50. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW)
821 version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes.
822 *Nucleic Acids Research* **47**, W59–W64 (2019).
823 51. Levy Karin, E., Mirdita, M. & Söding, J. MetaEuk—sensitive, high-throughput
824 gene discovery, and annotation for large-scale eukaryotic metagenomics.
825 *Microbiome* **8**, 48 (2020).
826 52. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**,
827 15–21 (2013).
828 53. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ
829 preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
830 54. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome
831 from RNA-seq reads. *Nat Biotechnol* **33**, 290–295 (2015).
832 55. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools.
833 *Bioinformatics* **25**, 2078–2079 (2009).
834 56. Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare. Preprint at
835 https://doi.org/10.12688/f1000research.23297.1 (2020).
836 57. Haas, B. Transdecoder.
837 58. Buchfink, B., Reuter, K. & Drost, H. G. Sensitive protein alignments at tree-of-
838 life scale using DIAMOND. *Nature Methods* **18**, 366–368 (2021).
839 59. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon
840 provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**,
841 417–419 (2017).

842 60. Baril, T., Galbraith, J. & Hayward, A. Earl Grey: A Fully Automated User-
843 Friendly Transposable Element Annotation and Analysis Pipeline. *Mol Biol Evol* **41**,
844 msae068 (2024).
845 61. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize implements and
846 enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
847 62. Alexa, A. & Rahnenfuhrer, J. topGO: enrichment analysis for gene ontology.
848 63. Gel, B. *et al.* regioneR: an R/Bioconductor package for the association
849 analysis of genomic regions based on permutation tests. *Bioinformatics* **32**, 289–291
850 (2016).
851 64. Gu, Z. & Hübschmann, D. SimplifyEnrichment: A Bioconductor Package for
852 Clustering and Visualizing Functional Enrichment Results. *genom. proteom.*
853 *bioinform.* **21**, 190–202 (2023).
854 65. Oggenfuss, U., Badet, T. & Croll, D. A systematic screen for co-option of
855 transposable elements across the fungal kingdom. *Mobile DNA* **15**, 2 (2024).
856 66. Prieto, M., Montané, N., Aragón, G., Martínez, I. & Rodríguez-Arribas, C.
857 Cyanobacterial Variability in Lichen Cephalodia. *Journal of Fungi* **9**, (2023).
858 67. Jüriado, I., Kaasalainen, U., Liira, J. & Rikkinen, J. Peltigera (Ascomycota)
859 living in open and shady environments depend on different Nostoc photobionts. *Folia*
860 *Cryptogamica Estonica* **60**, 75–99 (2023).
861 68. Hodkinson, B. P. *et al.* Lichen-symbiotic cyanobacteria associated with
862 Peltigera have an alternative vanadium-dependent nitrogen fixation system.
863 *European Journal of Phycology* **49**, 11–19 (2014).
864 69. Nelson, J. M. *et al.* Complete Genomes of Symbiotic Cyanobacteria Clarify
865 the Evolution of Vanadium-Nitrogenase. *Genome Biology and Evolution* **11**, 1959–
866 1964 (2019).
867 70. Garfias-Gallegos, D. *et al.* Central metabolism and development are rewired
868 in lichenized cyanobacteria. *ISME J* **19**, wraf166 (2025).
869 71. Vigil-Stenman, T., Ininbergs, K., Bergman, B. & Ekman, M. High abundance
870 and expression of transposases in bacteria from the Baltic Sea. *ISME J* **11**, 2611–
871 2623 (2017).
872 72. Muszewska, A., Steczkiewicz, K., Stepniewska-Dziubinska, M. & Ginalski, K.
873 Transposable elements contribute to fungal genes and impact fungal lifestyle. *Sci*
874 *Rep* **9**, 4307 (2019).

875 73. Muszewska, A., Steczkiewicz, K., Stepniewska-Dziubinska, M. & Ginalski, K.
876 Cut-and-Paste Transposons in Fungi with Diverse Lifestyles. *Genome Biology and*
877 *Evolution* **9**, 3463–3477 (2017).