bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 **Protein kinase gene declines linearly with elevation: a shared genomic feature across**
2 **species and continents in lichenized fungi suggests role in climate adaptation**

3 Dominik Merges[1,2,3*], Francesco Dal Grande[1,2,4,6], Henrique Valim[1,2], Garima Singh[1,2,4], Imke
4 Schmitt [1,2,5]

5 [1] Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany

6 [2] LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt am
7 Main, Germany

8 [3] Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural
9 Sciences, Box 7026, Uppsala 750 07, Sweden

10 [4] Department of Biology, University of Padova, Padua, Italy
11
12 [5] Department of Biological Sciences, Institute of Ecology, Evolution and Diversity, Goethe
13 University Frankfurt, Frankfurt am Main, Germany

14 [6] National Biodiversity Future Center (NBFC), Piazza Marina 61, Palermo, Italy
15
16

17 *Corresponding author:

18 Dominik Merges

19 Present address: Department of Forest Mycology and Plant Pathology, Swedish University of
20 Agricultural Sciences, Box 7026, Uppsala 750 07, Sweden
21 mergesd01@gmail.com
22
23
24 **Running title:** Gene presence/absence in lichens

25 **Keywords**: altitudinal gradient, environmental change, gene presences/absence, gene
26 loss/gain

27

28

29

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

30 **Abstract**

31 Intraspecific genomic variability affects a species' adaptive potential towards climatic
32 conditions. Variation in gene content across populations and environments may point at
33 genomic adaptations to specific environments. The lichen symbiosis, a stable association of
34 fungal and photobiont partners, offers an excellent system to study environmentally driven
35 gene content variation. Many species have remarkable environmental tolerances, and often
36 form populations in different climate zones. Here we combine comparative and population
37 genomics to assess the presence and absence of genes in high elevation and low elevation
38 genomes of two lichenized fungi of the genus *Umbilicaria*. The two species have non-
39 overlapping ranges, but occupy similar climatic niches in North America (*U. phaea*) and
40 Europe (*U. pustulata*): high elevation populations are located in the cold temperate zone and
41 low elevation populations in the Mediterranean zone. We assessed gene content variation
42 along replicated elevation gradients in each of the two species, based on a total of 2050
43 individuals across 26 populations. Specifically, we assessed shared orthologs across species
44 within the same climate zone, and tracked which genes increase or decrease in abundance
45 within populations along elevation. In total, we found 16 orthogroups with shared orthologous
46 genes in genomes at low elevation and 13 at high elevation. Coverage analysis revealed one
47 ortholog that is exclusive to genomes at low elevation. Conserved domain search revealed
48 domains common to the protein kinases (PKs) superfamily. We traced the discovered
49 ortholog in populations along five replicated elevation gradients on both continents. The
50 protein kinase gene linearly declined in abundance with increasing elevation, and was absent
51 in the highest populations. We consider the parallel loss of an ortholog in two species and in
52 two geographic settings a rare find, and a step forward in understanding the genomic
53 underpinnings of climatic tolerances in lichenized fungi. In addition, the tracking of gene
54 content variation provides a widely applicable framework for retrieving biogeographical
55 determinants of gene presence/absence patterns. Our work provides insights into gene content

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

56 variation of lichenized fungi in relation to climatic gradients, suggesting a new research
57 direction with implications for understanding evolutionary trajectories of complex symbioses
58 in relation to climatic change.

59

60

61
62

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

63 **Introduction**

64 The intraspecific genomic variability substantially affects a species' adaptive potential to
65 ecological interactions and climatic conditions (Badet et al., 2020; Hirsch et al., 2014; Resl et
66 al., 2022). Variation in genomic content at the gene-level, i.e. presence/absences of genes,
67 within species are regularly found in bacteria, and are often associated with adaptations to
68 specific environments and antibiotic resistances (Heuer et al., 2011; Masignani et al., 2005).
69 Recent studies suggest that eukaryotic species, like bacterial ones, can have intra-specific
70 variation in genomic content at the gene-level (Badet et al., 2020; Gerdol et al., 2020; Hirsch
71 et al., 2014). To this date, it is largely unknown how relevant such variation in gene content is
72 for eukaryotes, and if it is ecologically important for adaptations to specific environments.
73 Intra-specific variation in gene content contributes to the formation of genetically diverse
74 populations, and therefore its characterization is vital to understand the mechanisms of
75 adaptation to specific environments (Drott et al., 2021; Plissonneau et al., 2018).

76 The lichen symbiosis, a stable association of mainly fungal and photobiont partners as
77 well as an associated microbiome, lends itself to the study of environmentally driven gene
78 content, because many species have remarkable environmental tolerances and maintain
79 populations in different climate zones (Grimm et al., 2021; Jung et al., 2021; Kappen, 2000;
80 Medeiros et al., 2021; Sierra et al., 2020; Singh et al., 2017; Tanunchai et al., 2022; Werth &
81 Sork, 2014). Population genomic analyses based on single nucleotide polymorphisms (SNPs)
82 suggest the presence of genome-wide differentiation between populations in different climate
83 zones (Dal Grande et al., 2017; Rolshausen et al., 2022). Differentiation at the level of SNPs
84 is often significantly correlated with differences in geography and ecology and may thus be
85 involved in environmental specialization (Castillo et al., 2012; Hodkinson et al., 2012; Peksa
86 & Skaloud, 2011). The only study which assesses variation in gene content associated with
87 environmental adaptation is limited to a single species: Singh et al. (2021) show that some

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

88 gene clusters associated with natural product biosynthesis in *Umbilicaria pustulata* have
89 elevation-specific distributions. It is currently unknown 1) whether different species of lichen-
90 forming fungi maintain homologous population-specific genes along environmental gradients
91 and 2) whether intra-specific variation in gene content contributes to environmental tolerance
92 of lichens.

93 Modern evolutionary approaches leverage DNA sequencing to infer ecological and
94 evolutionary processes that occur at the population level. Here we combine comparative
95 genomics and population genomics to assess the presence/absence of genes in high elevation
96 and low elevation genomes of two lichenized fungi of the genus *Umbilicaria* (*Umbilicaria*
97 *phaea* and *U. pustulata*). The two species have non-overlapping ranges but occupy similar
98 climatic niches in North America (*U. phaea*) and Europe (*U. pustulata*), i.e. the cold
99 temperate and the Mediterranean zones. We tracked gene content variation along replicated
100 elevation gradients in both species based on a total of 2050 individuals in 26 populations.
101 Specifically, we addressed the following research questions: a) Which genes are linked to
102 environmental conditions at high elevation (cold temperate climate) and low elevation
103 (Mediterranean climate) across species? To address this question, we assessed which genes
104 are exclusive to climate zones. b) Do abundances of genes specific to a climate zone co-vary
105 with elevation in populations of *U. phaea* and *U. pustulata*? To address the second question,
106 we assessed which genes increase or decline in abundance with increasing elevation.

107

108 **Material and methods**

109 **Study site and sample collection**

110 We sampled 15 *U. pustulata* populations along three elevational gradients in Spain and Italy
111 and 11 *U. phaea* populations along two elevational gradients in California, USA
112 (Merges et al., 2021; Singh et al., 2022). Details of the sampling of European material are
113 described in Singh at al. (2022) and the details of sampling of the North American material in

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

114 (Dal Grande et al., 2017; Merges et al., 2021). Briefly, two of the European gradients are
115 located in Central Spain, Sierra de Gredos, and one on the island of Sardinia. We collected
116 fragments of 100 individuals each, at Mount Limbara (Sardinia, Italy; 6 populations, IT),
117 Sierra de Gredos (Sistema Central, Spain; 6 populations, ESii) and Talavera-Puerto de Pico
118 (Sistema Central, Spain; 3 populations, ESi), as described in Dal Grande et al. (2017). The
119 Californian gradients are spatially separated by approx. 700 km. We collected fragments of 50
120 individuals each, at four populations along the Sierra Nevada gradient (38.084, -129 120.484)
121 and at seven population along the Mt. Jacinto gradient (33.435, -116.484). We additionally
122 collected four whole lichen thalli, one low-altitude individual from the Sierra Nevada
123 population Uph16 and one from the high-altitude population Uph19, as well as a low-altitude
124 and a high-altitude individual from populations ESii1and Esii6 of the Sierra des Gredos
125 gradient for the reconstruction of reference genomes using PacBio Sequel II data.

126 **DNA extraction for population pooled sequencing**

127 Genomic DNA was extracted separately from each fragment from all populations using a
128 CTAB-based method (Cubero & Crespo, 2002; Dal Grande et al., 2017). Further, we created
129 a pooled sample for each population containing equal amounts of DNA from each sample
130 (i.e., Pool-seq; Dal Grande et al., 2017). Novogene Co., Ltd. (Cambridge, United Kingdom)
131 and performed the library preparation (200–300 bp insert size). Libraries were sequenced on
132 an Illumina HiSeq2000 with 150 bp paired-end chemistry at ~90x coverage per population.

133 **DNA extraction for genomic sequencing**

134 Genomic DNA for genome sequencing was extracted from dry thallus material of two
135 samples of the same species (i.e. *U. phaea* or *U. pustulata*) collected in different climatic
136 zones (i.e., low elevation/Mediterranean climate zone and high elevation/temperate climate
137 zone). Lichen thalli were thoroughly washed with sterile water and checked under the

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

138 stereomicroscope for the presence of possible contamination or other lichen thalli. DNA was
139 extracted from all of the samples using a cetyltrimethylammonium bromide (CTAB)-based
140 method (Mayjonade et al., 2016) as presented in Merges et al. (2021).

141 **PacBio library preparation and sequencing**

142 SMRTbell libraries were constructed according to the manufacturer's instructions of the
143 SMRTbell Express Prep kit v2.0 following the Low DNA Input Protocol (Pacific Biosciences,
144 Menlo Park, CA). Total input DNA was approximately 140 ng and 800 ng, respectively.
145 Ligation with T-overhang SMRTbell adapters was performed at 20°C overnight. Following
146 ligation, the SMRTbell library was purified with an AMPure PB bead clean up step with
147 0.45X volume of AMPure PB beads. Subsequently a size-selection step with AMPure PB
148 Beads was performed to remove short SMRTbell templates < 3kb. For this purpose, the
149 AMPure PB beads stock solution was diluted with elution buffer (40% volume/volume) and
150 then added to the DNA sample with 2.2X volume. The size and concentration of the final
151 libraries were assessed using the TapeStation (Agilent Technologies) and the Qubit
152 Fluorometer with Qubit dsDNA HS reagents Assay kit (Thermo Fisher Scientific, Waltham,
153 MA). Sequencing primer v4 and Sequel® II Polymerase 2.0 were annealed and bound,
154 respectively, to each SMRTbell library. SMRT sequencing was performed on the Sequel
155 System II with Sequel II Sequencing Kit 2.0 in "continuous long read" (i.e. CLR) mode, 30
156 hour movie time with no pre-extension and Software SMRTLINK 8.0. One SMRT Cell was
157 run for each sample.

158 **De novo assembly of PacBio metagenomic sequence reads**

159 We largely followed the pipeline described in Merges et al. (2021). In summary, we generated
160 HiFi reads from the Pacbio Sequel II run using the PacBio tool CCS v5.0.0 with default
161 parameters (https://ccs.how). Metagenomic sequence reads were assembled into contigs using

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

162 the long-read based assembler metaFlye v2.7 (Kolmogorov et al., 2019). The assembled
163 contigs were scaffolded with LRScaf v1.1.12 (https://github.com/shingocat/lrscaf; (Qin et al.,
164 2019)). To retrieve the mycobiont genome, the received scaffolds were taxonomically binned
165 via blastx using DIAMOND (--more-sensitive --frameshift 15 –range-culling) on a custom
166 database (Singh et al., 2022) and the MEGAN6 Community Edition pipeline (Buchfink et al.,
167 2014; Huson et al., 2007). The completeness of the genomes represented by the binned
168 Ascomycota scaffolds was estimated using Benchmarking Universal Single-Copy Orthologs
169 (BUSCO) analysis in BUSCO v4 (Simão et al., 2015) using the Ascomycota dataset.

170 **Pool-seq data processing**

171 We filtered the pool-seq data for reads shorter than 80 bp, reads with N's, and reads with
172 average base quality scores less than 26 along with their pairs, and discarded them. We
173 mapped the trimmed paired-end reads of each pool to the database of the identified genes
174 using bowtie2 v2.4.1 (Langmead & Salzberg, 2012), using the flags: --very-sensitive-local, --
175 no-mixed, --no-unal, --no-discordant.

176 **Gene prediction and genome annotation**

177 Functional annotation of genomes, including genes and proteins (antiSMASH; antibiotics &
178 SM Analysis Shell, v5.0) was performed with scripts implemented in the funannotate pipeline
179 (Blin et al., 2017; Palmer & Stajich, 2019). First, the genomes were masked for repetitive
180 elements, and then the gene prediction was performed using BUSCO2 to train Augustus and
181 self-training GeneMark- ES (Borodovsky & Lomsadze, 2011; Simão et al., 2015). Functional
182 annotation was done with InterProScan (Quevillon et al., 2005), egg-NOG-mapper (Huerta-
183 Cepas et al. 2019, 2017) (Huerta-Cepas et al., 2017, 2019), and BUSCO v 5.1.2 (Simão et al.,
184 2015) with ascomycota_db models. Secreted proteins were predicted using SignalP

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

185 (Armenteros et al., 2019) as implemented in the funannotate "annotate" command. Proteins
186 where further characterized by NCBI conserved domain search (Lu et al., 2020).

187 ***Assessing gene content variation in the assembled fungi genomes***

188 To identify presence/absences patterns of genes, we identified orthologs using orthoFinder
189 (Emms & Kelly, 2015, 2019). OrthoFinder provides the most accurate ortholog inference
190 method on the Quest for Orthologs benchmark test (Emms & Kelly, 2015, 2019). In
191 orthoFinder (v.2.5.4), we assigned all genes to orthogroups using protein homology and
192 constructed a pangenome of all four complete genomes (Badet et al., 2020). Shared orthologs
193 (i.e. members of the some orthogroup) of low elevation (warm adapted) and high elevational
194 (cold adapted) genomes were extracted using R v3.6.1 (R Core Team, 2019).

195 **Validating presence/absence of genes at population level**

196 To validate population level gene presence or absence, we estimated the abundance of each
197 ortholog in the low elevation (warm adapted) and high elevation (cold adapted) population
198 based on the median coverage of pool-seq reads associated to each ortholog contig.
199 Specifically, we used samtools (v1.15) depth to estimate the coverage of each basepair within
200 the contig (Danecek et al., 2021). We assessed and visualized the data in R v3.6.1 (R Core
201 Team, 2019).

202 **Gene distribution across Umbilicaria populations**

203 Bowtie2 (v2.2.2) was used to map pool-seq reads to all ortholog contigs (using default
204 settings). The number of mapped reads was counted per sample and normalized by dividing
205 the number of mapped reads by the total read number of the respective sample to account for
206 differences in sequencing depth. We modelled gene abundance (i.e., normalized read count)

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

207 as a function of elevation using linear models. Linear models were fitted and plotted in R
208 v3.6.1 (R Core Team, 2019).
209
210 **Results**
211 **HiFi metagenomic sequencing reads of mycobiont**
212 We reconstructed metagenomic sequences from a low-elevation and a high-elevation
213 specimen of *U. pustulata* and *U. phaea*. Sequence output and quality for *U. pustulata* were
214 summarized in Singh et al. 2022, for *U phaea* in Merges et al 2021 and in Table S1.
215 **Altitude-specific genes in the de-novo assembled genomes**
216 We screened the de-novo assembled genomes of the low- and high-altitude samples for
217 altitude-specific genes (Figure 1). Orthofinder revealed 16 orthogroups with shared
218 orthologous genes (0.2 % of the total orthogroups) in low-elevation genomes and 13 in high-
219 elevational genomes (0.1 % of the total orthogroups).
220 **Presence/absence of genes at population level**
221 To verify the presence/absence of detected orthologs, the coverage of each ortholog was
222 calculated for the respective population at low and high elevation. The coverage analysis
223 revealed one ortholog present in the genomes at low elevation to be consistently absent in
224 populations at high elevation (Fig. 2.). The amino acid sequences of the ortholog could not be
225 functionally annotated using the funannotate pipeline and was classified as "hypothetical
226 protein". NCBI's conserved domain search revealed an alignment with the catalytic domain
227 of Protein Kinases superfamily member PKc cd00180 (accession cl214531) as well as seven
228 Tetratricopeptide repeats, indicating putative protein binding surfaces.
229
230 **Gene abundance distributions along gradients**

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

231 The normalized read number of the identified orthologs, annotated as members of the Protein
232 Kinases superfamily, showed a decline with increasing elevation across all populations
233 (Figure 3).

234 **Discussion**

235 Although adaptations to environmental gradients may lead to variation in gene content,
236 assessments of gene presence/absence patterns across populations and species of lichenized
237 fungi are still missing. In this study, we take an in-depth look at gene presence and absences
238 in lichenized fungi of the genus *Umbilicaria*, and trace the discovered genes in lichen
239 populations along five replicated elevation gradients across two continents. While our whole
240 genome comparison based on four de novo sequenced specimen (two per species) suggested
241 up to 29 elevation-specific orthogroups, the population-level verification approach showed
242 only one gene, putatively encoding a protein kinase (PKs), which linearly declined in
243 abundance with increasing elevation, and was truly absent in the highest population. This
244 suggests either high strain-specificity of certain genes, or high false positive recovery of gene
245 presence/absence patterns when relying on comparative genomics approaches based on only a
246 few individuals. Such approaches can be potentially misleading when interpreting the
247 evolutionary significance of gene content variation at population level. Regarding the PKs
248 gene consistently absent in high elevation genomes and populations, we found that the
249 discovered gene declines linearly across all populations, suggesting an evolutionary benefit
250 only at lower altitudes. Alternatively, the loss of the gene at higher elevations might benefit
251 individuals in cold climates. To our knowledge, we report for the first-time parallel gene
252 presences and absences patterns correlating with climatic niches in different species of
253 lichenized fungi. However, it remains unclear if this is an adaptive trait.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

254 In bacteria variation in gene content is assumed to be driven by selection for
255 environmental conditions that are relatively rare across the entire range of a species (Qi et al.,
256 2017). Recent evidence suggests that specific populations of lichenized fungi may contain
257 unique biosynthetic gene clusters (Singh et al. 2021), and our current findings show that also
258 other genes can be elevation-specific. The gradual gene loss across populations with
259 increasing elevation may suggests a decline of selective benefit and may indicate that certain
260 variations in gene content could be of functional importance for local adaptation and climatic
261 tolerances in lichenized fungi. The conserved domain search revealed a catalytic domain of a
262 PK, a common eukaryotic protein superfamily. PKs selectively modify other proteins by
263 phosphorylation, changing their enzymatic activity, cellular location and association with
264 other proteins (Asano et al., 2005; Cheng et al., 2002; Hanks, 2003; Heinisch & Rodicio,
265 2018). Within a genome, PKs are encoded by a large multigene family with genes being
266 distributed among multiple chromosomes. Putatively, the high number of PK genes has arisen
267 by genome segmental duplication events (Asano et al., 2005; Heinisch & Rodicio, 2018). In
268 our study, the presence/absence of a single PK gene may suggest a climate-specific ancestral
269 genome segmental duplication event. However, due to the scarcity of functional annotation of
270 non-model organisms and the resulting lack of in-depth functional annotation of the gene in
271 question, the mechanisms generating such population level diversification are yet to be
272 understood.

273 While environmental adaptations are commonly highly polygenic (Barghi et al., 2019;
274 Hartke et al., 2021; Pfenninger et al., 2021; Rivas et al., 2018), there is increasing evidence of
275 the effect of single gene content variation (Liu et al., 2021). For example, as has been recently
276 shown in agave, where a single gene encoding a phosphoenolpyruvate carboxylase enhances
277 the plant's climate resilience (Liu et al., 2021). Not only the gain of genes, but also the loss of
278 genes has been associated with adaptive traits, such as the evolution of particular diets in bats

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

279 (Blumer et al., 2022). Therefore, we consider the parallel loss of a homologous gene in two
280 species and two geographic settings a rare find, and a step forward in understanding the
281 genomic underpinnings of climatic tolerances in lichenized fungi. Future research should
282 address the functional importance of the gene present at low altitude in the Mediterranean
283 climate zone, and specifically explore the effects of variation in gene abundances across
284 populations. Additionally, future research should consider using heterologous expression
285 approaches to reveal whether the gene presence could induce tolerances to warm conditions.
286
287 **Conclusion**

288 Our study demonstrates how comparative genomics in combination with population genomic
289 data can detect patterns of gene content variation across climatic gradients. In addition, the
290 tracking of gene content variation across populations provides a widely applicable framework
291 for retrieving meaningful biogeographical determinants of gene presence/absence patterns. To
292 this end our work provides insights into gene content variation of lichenized fungi in relation
293 to climatic gradients. This suggests a promising new research direction with implications for
294 understanding evolutionary trajectories in relation to climatic change.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

295 **Data Accessibility**
296 Raw sequence reads were deposited in the Sequence Read Archive under the BioProject
297 PRJNA693984 and PRJNA820300.
298
299 **Acknowledgments**
300 We thank Carola Greve and Jürgen Otte for the laboratory procedures, and Christoph Sinai,
301 Tilman Schell and Anjuli Calchera (Frankfurt am Main) for support with bioinformatics. We
302 thank Barbara Feldmeyer and Markus Pfenninger (Frankfurt am Main) for their valuable
303 suggestions on analyses.
304
305 **Funding information**
306 This research has been partly funded by the Hesse's "state initiative for the development of
307 scientific and economic excellence" (LOEWE initiative).
308
309
310 **Authors' contributions**
311 D.M., and I.S. conceived the ideas; I.S. and F.D.G. collected the data, D.M., G.S., H.M.
312 performed genome assembly and annotations, D.M. analyzed data, F.D.G. provided analytical
313 guidance; D.M. and I.S. wrote the manuscript. All authors contributed to the various drafts
314 and gave final approval for publication.
315

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

316 **References**

317 Armenteros, A., Juan, J., Konstantinos, D., Kaae, C., Version, D., & Armenteros, A. (2019).
318 *SignalP 5 . 0 improves signal peptide predictions using deep neural networks*.
319 https://doi.org/10.1038/s41587-019-0036-z
320 Asano, T., Tanaka, N., Yang, G., Hayashi, N., & Komatsu, S. (2005). *Genome-wide*
321 *Identification of the Rice Calcium-dependent Protein Kinase and its Closely Related*
322 *Kinase Gene Families : Comprehensive Analysis of the CDPKs Gene Family in Rice*.
323 *46*(2), 356–366. https://doi.org/10.1093/pcp/pci035
324 Badet, T., Oggenfuss, U., Abraham, L., Mcdonald, B. A., & Croll, D. (2020). *A 19-isolate*
325 *reference-quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici*.
326 1–18.
327 Barghi, N., Tobler, R., Nolte, V., Jaksic, A. M., Mallard, F., Otte, K. A., Dolezal, M., Taus,
328 T., Kofler, R., & Schlötterer, C. (2019). Genetic redundancy fuels polygenic adaptation
329 in Drosophila. *Plos Biology*, *17*(2), 1–31.
330 Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Duran, H. G. S.,
331 Santos, E. L. C. D. L., Kim, U., Nave, M., Dickschat, J. S., Mitchell, D. A., Shelest, E.,
332 Breitling, R., Takano, E., Lee, S. Y., Weber, T., & Medema, H. (2017). *antiSMASH 4 . 0*
333 *— improvements in chemistry prediction and gene cluster boundary identification*.
334 *45*(April), 36–41. https://doi.org/10.1093/nar/gkx319
335 Blumer, M., Brown, T., Freitas, M. B., Destro, A. L., Oliveira, J. A., Morales, A. E., Schell,
336 T., Greve, C., Pippel, M., Jebb, D., Hecker, N., Ahmed, A., Kirilenko, B. M., Foote, M.,
337 Janke, A., Lim, B. K., & Hiller, M. (2022). *Gene losses in the common vampire bat*
338 *illuminate molecular adaptations to blood feeding*. *6494*(March).
339 Borodovsky, M., & Lomsadze, A. (2011). Eukaryotic Gene Prediction Using
340 GeneMark.hmm-E and GeneMark-ES. *Curr Protoc Bioinformatics*, 1–11.
341 https://doi.org/10.1002/0471250953.bi0406s35.
342 Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using
343 DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176
344 Castillo, R. V., Beck, A., & Lumbsch, H. T. (2012). Photobiont selectivity and specificity in
345 Caloplaca species in a fog-induced community in the Atacama Desert , northern Chile.
346 *Fungal Biology*, *116*(6), 665–676. https://doi.org/10.1016/j.funbio.2012.04.001
347 Cheng, S., Willmann, M. R., Chen, H., & Sheen, J. (2002). *Update on Calcium Signaling*
348 *Calcium Signaling through Protein Kinases . The Arabidopsis Calcium-Dependent*
349 *Protein Kinase Gene Family 1*. *129*(June), 469–485. https://doi.org/10.1104/pp.005645.1
350 Dal Grande, F., Sharma, R., Meiser, A., Rolshausen, G., Büdel, B., Mishra, B., Thines, M.,
351 Otte, J., Pfenninger, M., & Schmitt, I. (2017). Adaptive differentiation coincides with
352 local bioclimatic conditions along an elevational cline in populations of a lichen-forming
353 fungus. *BMC Evolutionary Biology*, *17*(93), 1–14. https://doi.org/10.1186/s12862-017-
354 0929-8
355 Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A.,
356 Keane, T., Mccarthy, S. A., Davies, R. M., & Li, H. (2021). *Twelve years of SAMtools*
357 *and BCFtools*. *December 2020*, 1–4. https://doi.org/10.1093/gigascience/giab008
358 Drott, M. T., Rush, T. A., Satterlee, T. R., Giannone, R. J., & Abraham, P. E. (2021).
359 *Microevolution in the pansecondary metabolome of Aspergillus flavus and its potential*
360 *macroevolutionary implications for filamentous fungi*. 1–10.
361 https://doi.org/10.1073/pnas.2021683118
362 Emms, D. M., & Kelly, S. (2015). OrthoFinder : solving fundamental biases in whole genome
363 comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 1–
364 14. https://doi.org/10.1186/s13059-015-0721-2
365 Emms, D. M., & Kelly, S. (2019). *OrthoFinder : phylogenetic orthology inference for*
366 *comparative genomics*. 1–14.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

367 Gerdol, M., Moreira, R., Cruz, F., Gómez-garrido, J., Vlasova, A., Rosani, U., Venier, P.,
368 Naranjo-ortiz, M. A., Murgarella, M., Greco, S., Balseiro, P., & Corvelo, A. (2020).
369 Massive gene presence-absence variation shapes an open pan-genome in the
370 Mediterranean mussel. *Genome Biology*, *21*(275), 1–21.
371 Grimm, M., Grube, M., Schiefelbein, U., Zühlke, D., & Bernhardt, J. (2021). *The Lichens '*
372 *Microbiota , Still a Mystery ? 12*(March), 1–25.
373 https://doi.org/10.3389/fmicb.2021.623839
374 Hanks, S. K. (2003). *Genomic analysis of the eukaryotic protein kinase superfamily :*
375 Hartke, J., Waldvogel, A., Sprenger, P. P., Schmitt, T., Menzel, F., Pfenninger, M., &
376 Feldmeyer, B. (2021). *Little parallelism in genomic signatures of local adaptation in two*
377 *sympatric , cryptic sister species*. *July 2020*, 937–952. https://doi.org/10.1111/jeb.13742
378 Heinisch, J., & Rodicio, R. (2018). *Protein kinase C in fungi — more than just cell wall*
379 *integrity J urgen*. *October 2017*, 22–39. https://doi.org/10.1093/femsre/fux051
380 Heuer, H., Schmitt, H., & Smalla, K. (2011). *Antibiotic resistance gene spread due to manure*
381 *application on agricultural fields*. 236–243. https://doi.org/10.1016/j.mib.2011.04.009
382 Hirsch, C. N., Foerster, J. M., Johnson, J. M., Sekhon, R. S., Muttoni, G., Vaillancourt, B.,
383 Peñagaricano, F., Lindquist, E., Pedraza, A., Barry, K., Leon, N. De, Kaeppler, S. M., &
384 Buell, C. R. (2014). *Insights into the Maize Pan-Genome and Pan-Transcriptome*.
385 *26*(January), 121–135. https://doi.org/10.1105/tpc.113.119982
386 Hodkinson, B. P., Gottel, N. R., Schadt, C. W., & Lutzoni, F. (2012). *Photoautotrophic*
387 *symbiont and geography are major factors affecting highly structured and diverse*
388 *bacterial communities in the lichen microbiome*. *14*, 147–161.
389 https://doi.org/10.1111/j.1462-2920.2011.02560.x
390 Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Mering, C. Von,
391 Bork, P., & Delbru, M. (2017). Fast Genome-Wide Functional Annotation through
392 Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.*, *34*(8), 2115–2122.
393 https://doi.org/10.1093/molbev/msx148
394 Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hern, A., Forslund, S. K., Cook, H., Mende, D.
395 R., Letunic, I., Rattei, T., Jensen, L. J., Mering, C. Von, & Bork, P. (2019). eggNOG 5 .
396 0 : a hierarchical , functionally and phylogenetically annotated orthology resource based
397 on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(November 2018), 309–
398 314. https://doi.org/10.1093/nar/gky1085
399 Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of
400 metagenomic data. *Genome Research*, *17*(3), 377–386.
401 https://doi.org/10.1101/gr.5969107
402 Jung, P., Brust, K., Schultz, M., Büdel, B., Donner, A., Lakatos, M., Ruprecht, U., Beck, A.,
403 & Orlando, J. (2021). *Opening the Gap : Rare Lichens With Rare Cyanobionts –*
404 *Unexpected Cyanobiont Diversity in Cyanobacterial Lichens of the Order Lichinales*.
405 *12*(October). https://doi.org/10.3389/fmicb.2021.728378
406 Kappen, L. (2000). *Some aspects of the great success of lichens in Antarctica*. *72*(3), 314–
407 324.
408 Kolmogorov, M., Rayko, M., Yuan, J., Polevikov, E., & Pevzner, P. (2019). metaFlye:
409 scalable long-read metagenome assembly using repeat graphs. *BioRxiv*, 637637.
410 https://doi.org/10.1101/637637
411 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
412 *Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923
413 Liu, D., Hu, R., Zhang, J., Guo, H., Cheng, H., Li, L., & Yang, X. (2021). *Overexpression of*
414 *an Agave Phospho enol pyruvate Carboxylase Improves Plant Growth and Stress*
415 *Tolerance*. 1–20.
416 Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, R., Gwadz, M.,
417 Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M.,

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

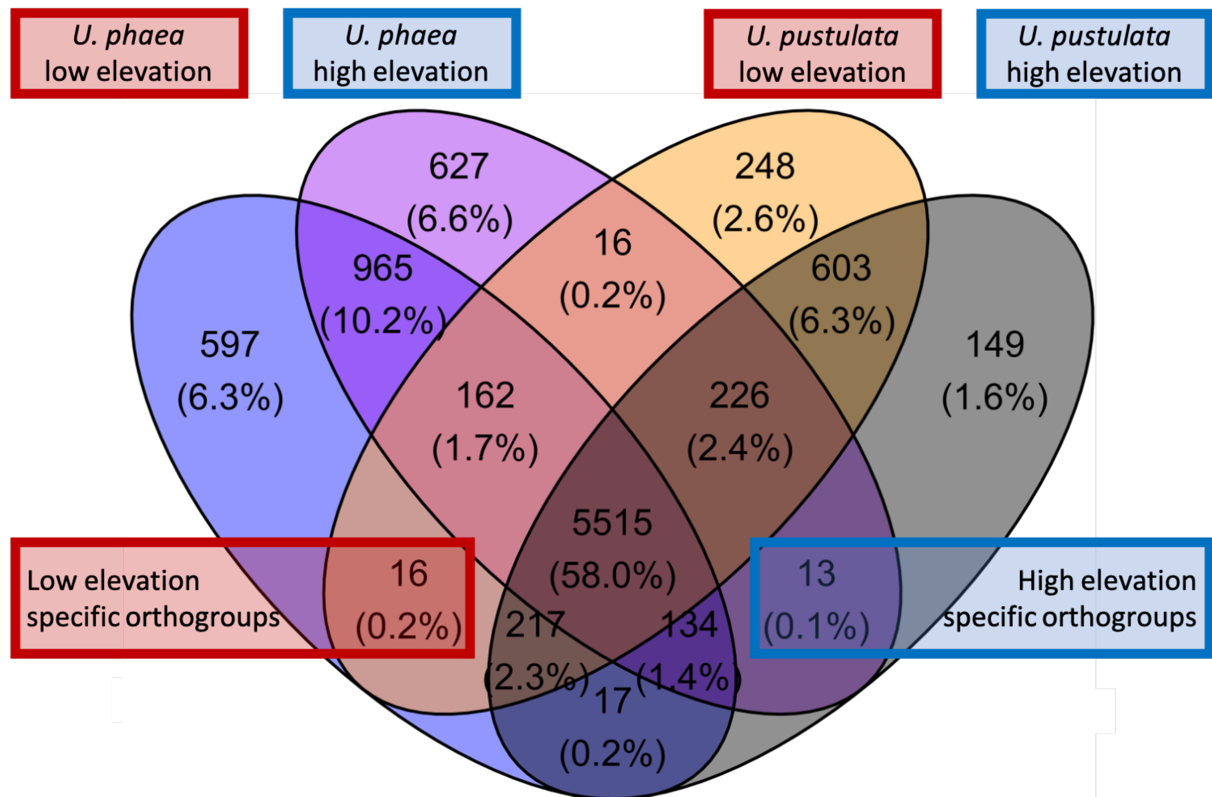
418 Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-bauer, A. (2020). *CDD / SPARCLE : the conserved domain database in 2020*. *48*(November 2019), 265–268.
420 https://doi.org/10.1093/nar/gkz991
421 Masignani, V., Medini, D., Donati, C., & Rappuoli, R. (2005). *The microbial pan-genome*.
422 589–594. https://doi.org/10.1016/j.gde.2005.09.006
423 Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Callot, C., Langlade, N., & Munos, S.
424 (2016). Extraction of high-molecular-weight genomic DNA for lang-read sequencing of
425 single molecules. *BioTechniques*, *61*, 203–205. https://doi.org/10.2144/000114460
426 Medeiros, I., Mazur, E., Miadlikowska, J., Flakus, A., Rodriguez-Flakus, P., Pardo-De la Hoz,
427 C., Cieslak, E., Sliwa, L., & Lutzoni, F. (2021). *Turnover of Lecanoroid Mycobionts and*
428 *Their Trebouxia Photobionts Along an Elevation Gradient in Bolivia Highlights the Role*
429 *of Environment in Structuring the Lichen Symbiosis*. *12*(774839).
430 https://doi.org/10.3389/fmicb.2021.774839
431 Merges, D., Grande, F. D., Greve, C., Otte, J., & Schmitt, I. (2021). *Virus diversity in*
432 *metagenomes of a lichen symbiosis (Umbilicaria phaea): complete viral genomes ,*
433 *putative hosts and elevational distributions*. *00*. https://doi.org/10.1111/1462-
434 2920.15802
435 Palmer, J., & Stajich, J. (2019). *Funannotate v1.7.4*. Zenodo.
436 Peksa, O., & Skaloud, P. (2011). Do photobionts influence the ecology of lichens ? A case
437 study of environmental preferences in symbiotic green alga Asterochloris
438 (Trebouxiophyceae). *Molecular Ecology*, *20*, 3936–3948. https://doi.org/10.1111/j.1365-
439 294X.2011.05168.x
440 Pfenninger, M., Reuss, F., Kiebler, A., Mody, K., Mishra, B., & Thines, M. (2021). *Genomic*
441 *basis for drought resistance in European beech forests threatened by climate change*. 1–
442 17.
443 Plissonneau, C., Hartmann, F. E., & Croll, D. (2018). *Pangenome analyses of the wheat*
444 *pathogen Zymoseptoria tritici reveal the structural basis of a highly plastic eukaryotic*
445 *genome*. 1–16. https://doi.org/10.1186/s12915-017-0457-4
446 Qi, Q., Zhao, M., Wang, S., Ma, X., Wang, Y., Gao, Y., Lin, Q., Li, X., Gu, B., Li, G., Zhou,
447 J., & Yang, Y. (2017). *The Biogeographic Pattern of Microbial Functional Genes along*
448 *an Altitudinal Gradient of the Tibetan Pasture*. *8*(976), 1–12.
449 https://doi.org/10.3389/fmicb.2017.00976
450 Qin, M., Wu, S., Li, A., Zhao, F., Feng, H., Ding, L., & Ruan, J. (2019). *LRScaf : improving*
451 *draft genomes using long noisy reads*. 1–12.
452 Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., Lopez, R.,
453 Bucher, P., Cerutti, L., & Res, A. (2005). *InterProScan : protein domains identifier*. *33*,
454 116–120. https://doi.org/10.1093/nar/gki442
455 Resl, P., Bujold, A. R., Tagirdzhanova, G., Meidl, P., Rallo, S. F., Kono, M., Fernández-
456 brime, S., Guðmundsson, H., Andrésson, Ó. S., Muggia, L., Mayrhofer, H., Mccutcheon,
457 J. P., Wedin, M., Werth, S., Willis, L. M., & Spribille, T. (2022). *transport potential*
458 *among lichen fungal symbionts*. 1–13. https://doi.org/10.1038/s41467-022-30218-6
459 Rivas, M. J., Saura, M., Pérez-figueroa, A., Panova, M., Johansson, T., André, C., Caballero,
460 A., Rolán-alvarez, E., Johannesson, K., & Quesada, H. (2018). *Population genomics of*
461 *parallel evolution in gene expression and gene sequence during ecological adaptation*.
462 *April*, 1–12. https://doi.org/10.1038/s41598-018-33897-8
463 Rolshausen, G., Dal, F., Jürgen, G., & Imke, O. (2022). *Lichen holobionts show*
464 *compositional structure along elevation*. *October 2021*, 1–12.
465 https://doi.org/10.1111/mec.16471
466 Sierra, M. A., Danko, D. C., Sandoval, T. A., Pishchany, G., Moncada, B., Kolter, R., Mason,
467 C. E., & Zambrano, M. M. (2020). *The Microbiomes of Seven Lichen Genera Reveal*
468 *Host Specificity , a Reduced Core Community and Potential as Source of Antimicrobials*.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

469 *11*(March), 1–12. https://doi.org/10.3389/fmicb.2020.00398
470 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015).
471 BUSCO: Assessing genome assembly and annotation completeness with single-copy
472 orthologs. *Bioinformatics*, *31*(19), 3210–3212.
473 https://doi.org/10.1093/bioinformatics/btv351
474 Singh, G., Calchera, A., Merges, D., Valim, H., Otte, J., & Schmitt, I. (2022). A Candidate
475 Gene Cluster for the Bioactive Natural Product Gyrophoric Acid in Lichen-Forming
476 Fungi. *Microbiology Spectrum*, *10*(4).
477 Singh, G., Grande, F. D., Divakar, P. K., Crespo, A., & Schmitt, I. (2017). Fungal – algal
478 association patterns in lichen symbiosis linked to macroclimate. *New Phytologist*,
479 *214*(1), 317–329.
480 Tanunchai, B., Schroeter, S. A., Ji, L., Fareed, S., Wahdan, M., Hossen, S., Lehnert, A.,
481 Grünberg, H., Gleixner, G., Buscot, F., Schulze, E., Noll, M., & Purahong, W. (2022).
482 More than you can see : Unraveling the ecology and biodiversity of lichenized fungi
483 associated with leaves and needles of temperate tree species using high-throughput
484 sequencing. *Frontiers in Microbiology*, *13*(907531).
485 Werth, S., & Sork, V. L. (2014). ECOLOGICAL SPECIALIZATION IN TREBOUXIA
486 (TREBOUXIOPHYCEAE) PHOTOBIONTS OF RAMALINA MENZIESII (
487 RAMALINACEAE) ACROSS SIX RANGE-COVERING ECOREGIONS OF
488 WESTERN NORTH AMERICA. *American Journal of Botany*, *101*(7), 1127–1140.
489 https://doi.org/10.3732/ajb.1400025
490
491

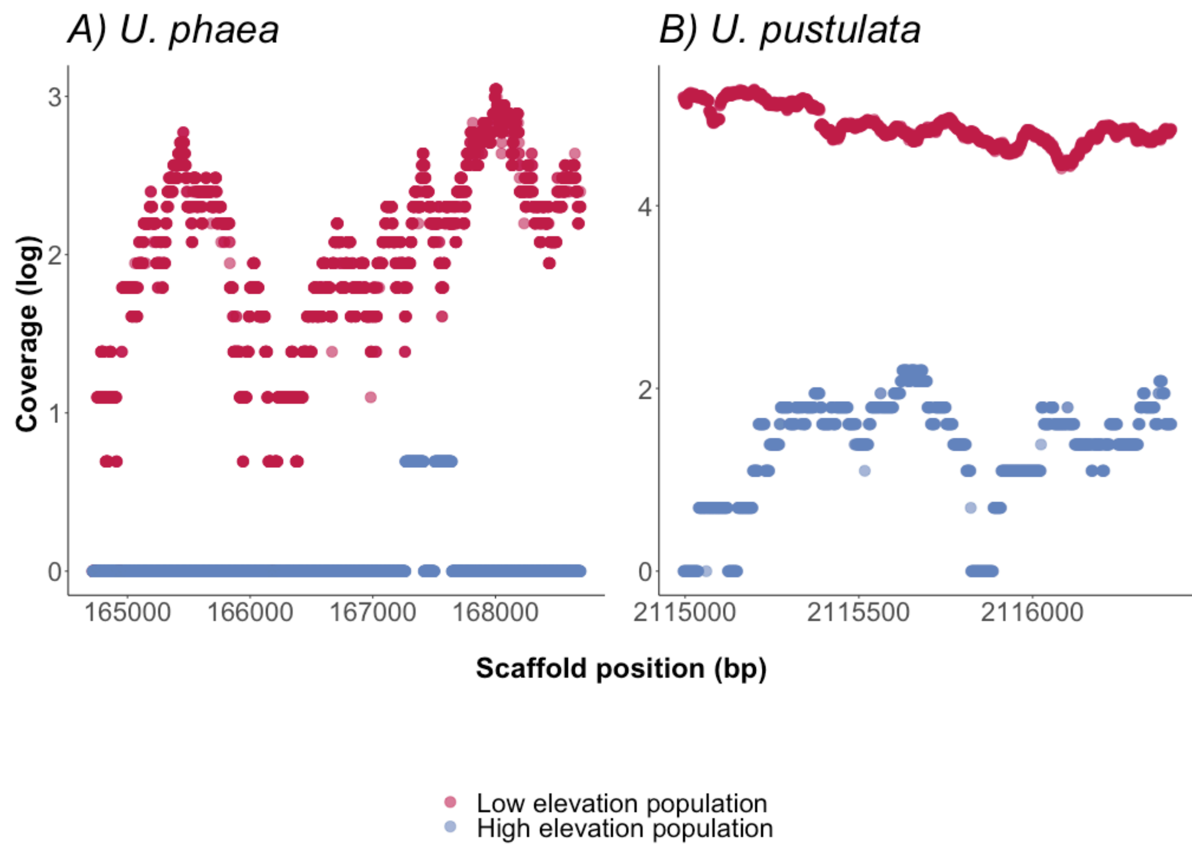
bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

492 **Figures**
493



494
495 Figure 1: Venn diagram displaying orthogroups of *U. phaea* and *U. pustulata*. Red box
496 highlights orthologs of the warm adapted (low elevation) *U. phaea* and *U. pustulata* genomes
497 and the blue box of the cold adapted (high elevation) genomes.
498

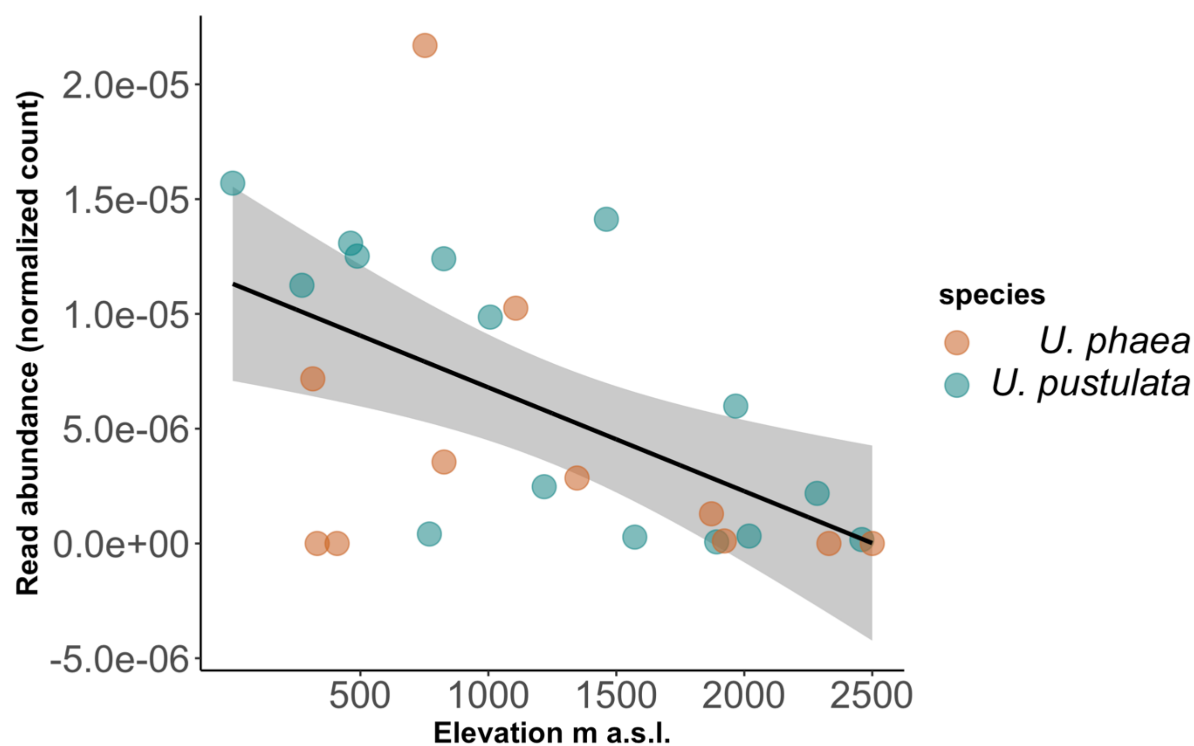
bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



499
500 Figure 2: Presence/absences of orthologs in four de novo sequenced genomes of *U. phaea* and
501 *U. pustulata* were verified by assessing the scaffold coverage in the warm adapted (low
502 elevation) and the cold adapted population (high elevation) respectively. A) Coverage of
503 ortholog in *U. phaea*: High coverage in warm adapted (low elevation) population and no
504 coverage in high elevational population. B) Coverage of ortholog in *U. pustulata*: High
505 coverage in in warm adapted (low elevation) population.
506

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.02.514805; this version posted November 2, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

507



508
509 Figure 3: Read abundance of identified ortholog decreases significantly with in increasing
510 elevation across *U. phaea* (brown circles) and *U. pustulata* (blue circles) populations.
511

512 Supplement
513

514 TABLE S1 Genome quality and annotation statistics

Taxon	Sample ID	CCS HiFi yield (%)	No. of scaffolds	N50	Completeness (%)	Assembly size (Mb)	No. of genes	No. of proteins	Genome accession
U. phaea 1	TBG_1111	5.72	47	1.54	96.5	35.1	7,659	7,576	SRX9925339
U. phaea 2	TBG_1112	22.36	38	1.22	96.5	35.55	7,681	7,628	SRX9925340
U. pustulata 1	TBG_2333	33	26	2.62	97.3	37.7	9,569	9,503	JALILU000000000
U. pustulata 2	TBG_2345	32.26	31	2.36	96.8	35.7	8,790	8,740	JALILV000000000

515

516