bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 **First genome-wide analysis of an endangered lichen reveals isolation by distance and**
2 **strong population structure**

3 **Running Title:** Rock Gnome Population Genomics

4 **Keywords:** Conservation Genetics, Fungi, Population Genetics, Genomics

5 Jessica L. Allen[*,a], Sean K. McKenzie[b], Robin S. Sleith[a], and S. Elizabeth Alter[c,d]

6 [a]The New York Botanical Garden, 2900 Southern Blvd., Bronx, NY 10458
7 [b]Rockefeller University, 1230 York Avenue, New York, NY 10065
8 [c]York College, Biology Department, 94-20 Guy R Brewer Blvd, Jamaica, NY 11451
9 [d]The Graduate Center, City University of New York, 365 5th Avenue, New York NY 10016
10 *corresponding author: jallen@nybg.org, (509) 412-2071

11 **Author Contributions**
12 Jessica Allen designed this research project, applied for and acquired all required permits and
13 funding for field work, conducted all field work and DNA extractions, analyzed data, and wrote
14 the manuscript. Sean McKenzie assisted with portions of the field work, advised on sequencing
15 techniques, and analyzed data. Robin Sleith advised on sequencing techniques and analyzed data.
16 Elizabeth Alter advised on research project design, data analysis, and assisted in writing the
17 manuscript.
18

19 **Acknowledgements**
20 We would like to thank the following people for aiding with this research: Alex Cecil and Jenna
21 Dorey for their field assistance, Gary Kauffman for help locating populations throughout the
22 National Forest and many helpful discussions about *Gymnoderma*, Chris Ulrey for help locating
23 the species on Blue Ridge National Park land, and especially for rapelling expertise, Dr. Richard
24 Harris for conducting microscopy to search for trichogynes and spores. Library preparations and
25 sequencing was conducted at The Rockefeller Genome Resource Center. Funding for this
26 research came from Highlands Biological Station, NSF GRFP, and NSF DEB#1145511.
27

28 **Data Accessibility** (To be deposited upon publication)
29 Reference genome: Genbank Accession XXXX
30 Raw reads: Sequence Read Archive XXXX
31

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

32 **First genome-wide analysis of an endangered lichen reveals isolation by distance and**
33 **strong population structure**

34 **Abstract**

35 Lichenized fungi are evolutionarily diverse and ecologically important, but little is known
36 about the processes driving diversification and genetic differentiation in these lineages. Though
37 few studies have examined population genetic patterns in lichens, their geographic distributions
38 are often assumed to be wholly shaped by ecological requirements rather than dispersal
39 limitations. Furthermore, while their reproductive structures are observable, the lack of
40 information about recombination mechanisms and rates can make inferences about reproductive
41 strategies difficult. Here we investigate the population genomics of *Cetradonia linearis*, an
42 endangered lichen narrowly endemic to the southern Appalachians of eastern North America, to
43 test the relative contributions of environmental factors and geographic distance in shaping
44 genetic structure, and to gain insights into the demography and reproductive biology of range
45 restricted fungi. Analysis of genome-wide SNP data indicated strong evidence for both low rates
46 of recombination and for strong isolation by distance, but did not support isolation by
47 environment. Hindcast species distribution models and the spatial distribution of genetic
48 diversity also suggested that *C. linearis* had a larger range during the last glacial maximum,
49 especially in the southern portion of its current extent, consistent with previous findings in other
50 southern Appalachian taxa. These results contribute to our understanding of intrinsic and
51 extrinsic factors shaping genetic diversity and biogeographic patterns in *C. linearis*, and more
52 broadly, in rare and endangered fungi.

53 **Keywords:** Conservation Genetics, Fungi, Population Genetics, Genomics

54

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

55 **Introduction**

56 Symbiotic fungi, including lichenized species, represent some of the most ecologically
57 important radiations on earth. However, the processes shaping genetic differentiation and gene
58 flow in these groups remain poorly understood. Historically, two major assumptions have shaped
59 hypotheses about symbiotic fungal population structure and evolution. First, because most fungi
60 produce very small spores, their distribution is thought to be limited primarily by ecological
61 suitability rather than geographic distance (O'Malley 2007). Second, species in which no sexual
62 reproductive structures have been observed are assumed to reproduce only asexually (Taylor et
63 al. 2015). Phylogenetic and population genetic studies have already challenged these
64 assumptions in fungi that do not form lichens. For instance, in the common and widespread
65 fungus *Suillis brevipes* there is evidence for both isolation by distance (IBD) and adaptation of
66 coastal populations to saline environmental gradients (Branco et al. 2015). Species in
67 *Saccharomyces* show varying levels of geographic structure in their genetic differentiation, with
68 *S. paradoxus* showing clear evidence of IBD and *S. cerevisiae* showing much less geographic
69 structure (Liti et al. 2009). Taylor et al. (2015) reviewed the literature on clonal reproduction in
70 fungi, concluding that numerous species showed evidence for recombination regardless of
71 observed reproductive structures. Additional examples of similar observations, indicating that
72 fungal reproduction and population genetics are more complex than previously expected, have
73 been derived from population genetic and genomic data in other groups of non-lichenized fungi
74 (see review in Grünwald et al. (2016) and Peter and Schacherer (2016)).

75 Lichens are a major group of fungi that form obligate symbioses with algae and/or
76 cyanobacteria and comprise >20% of all ascomycetes (Lücking et al. 2016). Despite their
77 ecological importance and conspicuous abundance in many terrestrial ecosystems, relatively few

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

78 taxa have been studied with traditional population genetics methods, and no published studies
79 have used a genome-wide approach to assess gene flow or other population-level attributes. To
80 date, most population genetics studies of lichens have been conducted on *Lobaria pulmonaria*
81 and its algal photobiont *Dictyochlorpsis reticulata* using microsatellite markers (Widmer et al.
82 2010, Dal Grande et al. 2010, Nadyeina et al. 2014). These studies have shown that *L.*
83 *pulmonaria* frequently disperses short distances via lichenized propagules (bundles of algae and
84 fungi), infrequently disperses long distances via sexually produced fungal spores (Werth et al.
85 2006), and that there is evidence of adaptation and population isolation on small spatial scales
86 (Nadyeina et al. 2014). Population genetic patterns in another lichenized fungus, *Xanthoria*
87 *parietina,* based on RAPD-PCR markers, contrast starkly with the findings for *L. pulmonaria;* in
88 the former, high genetic diversity and very few clones were found within small areas, even
89 among adjacent individuals (Itten and Honneger 2010). The pattern recovered in *X. parietina* is
90 similar to a study of *Parmelina carporrhizans* based on microsatellite loci, where high rates of
91 migration were recovered among populations, except for isolated island populations (Alors et al.
92 2017). A microsatellite-based study on *Parmotrema tinctorum* and its algal symbiont found that
93 most dispersal was clonal over short distances, similar to *L. pulmonaria*, but still found evidence
94 for high rates of sexual reproduction in the fungus (Mansournia et al. 2012). While highly
95 detailed, these studies of lichen population genetics represent only a fraction of this diverse
96 group of fungi that have evolved an obligate symbiotic lifestyle at least seven times
97 independently throughout the fungal tree of life (Schoch et al. 2009), and occupy every terrestrial
98 ecosystem from the poles to the tropics (Brodo et al. 2001). Microsatellites have recently been
99 developed for a broader diversity of lichenized fungi (Magain et al. 2010; Devkota et al. 2014;

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

100 Nadyeina et al. 2014; Lindgren et al. 2016; Lutsak et al. 2016), however these tools have not yet
101 been utilized for population genetic analyses in lichens.
102 Population genomics is a promising approach to rapidly advance our knowledge of
103 population biology in lichens as it circumvents difficulties associated with developing species-
104 specific markers, especially since lichens are notoriously difficult and slow to culture (Crittenden
105 *et al*. 1995). Of the domains of eukaryotic organisms, fungi are one of the most amenable to
106 genomic studies due to their generally small, compact genomes (Gladieux et al. 2014).
107 Population genomics studies have already added substantial depth and breadth to the knowledge
108 of the basic fungal biology, allowing researchers to address questions that were once intractable
109 about life history and evolution of reproductive systems. For example, fungi that have only been
110 observed reproducing asexually show genomic evidence for sexual reproduction (Tsai et al.
111 2008; Stefanini et al. 2016); speciation through homoploid hybridization has been shown to
112 occur rapidly, at least in yeast (Leducq et al. 2016); and Glomerales, arbuscular mycorrhizal
113 fungi, have highly flexible levels of ploidy in the heterokaryotic cells within species (Wyss et al.
114 2016). Applying these methods to lichenized fungi holds great promise to rapidly advance
115 knowledge of lichen population biology.
116 The rock gnome lichen (*Cetradonia linearis*) is one of two fungal species protected by
117 the Endangered Species Act in the United States (USFWS 2013), and one of eight lichens on the
118 IUCN Red-List (Allen et al. 2015). It is narrowly endemic to the Southern Appalachians of
119 eastern North America, where it is known from ~100 populations, most of which are located in
120 western North Carolina (USFWS 2013). It forms colonies on rocks either on exposed cliffs at
121 high-elevations or on large boulders in mid- to high-elevation streams. *Cetradonia* is a
122 monotypic genus, whose position as the earliest diverging member of the widespread and

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

123 ecologically important Cladoniaceae makes its study essential for addressing hypotheses of
124 evolution in this family (Wei and Ahti 2002; Zhou et al. 2006). It forms colonies of simple to
125 branched squamules with black apothecia and/or pycnidia, reproductive structures, frequently
126 produced at the tips (Fig.1). Despite having been protected by the Endangered Species Act for
127 over 20 years, little is known about *C. linearis* beyond its distribution (USFWS 2013). Currently,
128 nothing is known about dispersal or population genetic structure in this species.

129 In this study, we tested three hypotheses concerning population-level processes in
130 *Cetradonia linearis*: 1) most reproduction and dispersal occurs through clonal processes, 2)
131 isolation by distance is the major force shaping the genetic differentiation, while ecological
132 adaptation plays a minor role, and 3) the southern portion of its current extent represents a major
133 refugium during the Pleistocene glaciation. To test these hypotheses, low-coverage, whole
134 genome shotgun sequencing was used to generate large-quantities of genomic data from samples
135 throughout the species' range. The resulting genome-wide single-nucleotide polymorphisms
136 (SNPs) were used to measure genetic diversity, recombination, and clonality. Population genetic
137 structure, connectivity, and evidence for isolation by environment were also investigated. This
138 study is the first assessment of population genomics in a lichen, providing a baseline for
139 comparison in this group of organisms, along with valuable information for the continued
140 conservation of the endangered rock gnome lichen.

141 **Methods**

142 *Study System, Sampling, and Sequencing*

143 Samples were collected from 15 sites throughout the geographic and ecological range of
144 *Cetradonia linearis* (Fig. 2). At each site two to three squamules were taken from up to ten
145 distinct colonies using surface sterilized forceps. Squamules were placed into 1.5 mL Eppendorf

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

146 tubes, set out to air dry for 24 hours, then stored at -40°C. Samples were washed with acetone
147 and DNA was extracted using the Qiagen DNeasy Plant Mini Kit with the cell lysis stage
148 extended for 4-6 hours. Thirty-two samples were chosen for sequencing based on DNA quality
149 and yield, while maintaining the geographic and ecological breadth of samples. Sequencing was
150 conducted at the Rockefeller University Genomics Resource Center. Libraries were prepared
151 with the Nextera XT kit and sequenced on the Illumina Next Seq platform in Mid Output on 150
152 bp paired end read mode. All samples were sequenced at roughly equal coverage, except one
153 sample from the Balsam Mountains, B224, which was sequenced at 5x higher coverage for
154 assembly of a reference genome.

155 *Quality Filtering, Genome Assembly, and Annotation*

156 A reference genome was assembled from sample B224 and annotated after strictly
157 filtering contaminating reads (see below). B224 reads were trimmed, adapters were removed,
158 and overlapping read-pairs combined using cutadapt and FLASH (Magoč and Salzberg 2011;
159 Martin 2011). Read pools for all other samples were trimmed, adapters were removed, and
160 overlapping read-pairs combined using FLASH and Trimmomatic v 0.36 (Bolger et al. 2014).
161 An initial assembly of B224 was built using Minia with a kmer size of 75 and an abundance
162 minimum of 3 (Chikhi and Rizk 2013). To filter out contaminants (including algal symbionts)
163 the Blobology workflow and perl scripts were used (Kumar et al. 2013). Specifically, a random
164 subset of 15,000 contigs longer that 250 bp were subjected to homology search using megablast
165 against the non-redundant nucleotide database from Genbank and the e-value cutoff was set to
166 1e-5. Based on these plots (Supplementary Fig. 1) contigs with GC content >0.6 and coverage <5
167 were pooled to form a set of contaminant contigs. Then, all B224 reads were aligned to the
168 contaminant contigs using bowtie2, and all reads that did not align to the contaminants were

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

169 retained for reassembly. The final assembly was built using Abyss with the paired-end read
170 setting and a kmer size of 41 (Simpson et al. 2009). All resulting contigs shorter than 500 bp
171 were removed from the dataset before further analyses. Genome annotation was conducted using
172 the MAKER pipeline (Cantarel et al. 2008). SNAP was used for the ab-initio gene predictor, and
173 protein homology evidence was drawn from *Aspergillus niger* ATCC 1015 v4.0, *Cladonia grayi*
174 Cgr/DA2myc/ss v2.0, and *Cochliobolus heterostrophus* C5 v2.0 (Andersen et al. 2011; Ohm et
175 al. 2012; Condon et al. 2013; McDonald et al. 2013; Leskovec and Sosič 2016). For a final
176 filtering step, all genes were blasted against the *A. niger*, *C. grayi*, and *C. heterostrophus* gene
177 sets. Contigs were only kept for downstream analysis if the gene with the highest-scoring blast
178 hit matched most closely with a *C. grayi* gene. Because the *C. grayi* genome was assembled
179 from pure culture of the fungal symbiont, this was an additional step that filtered any remaining
180 contaminants from the genome. Finally, we conducted homology searches for both mating-type
181 idiomorphs (MAT1-2 and MAT2-2) in the genome and all sampled read pools (Supplementary
182 Text).

183 *SNP Calling and descriptive statistics*

184 Single-nucleotide polymorphisms were called for all sequenced samples using the
185 annotated contigs as a reference genome for the fungal component. First, we used bwa (Li and
186 Durbin 2009) to align the reads to the contigs. Then, to call the SNPs from this alignment
187 FreeBayes was used with the ploidy set to 2, minimum alternate fraction set to 0.9 (Garrison and
188 Marth 2012). The ploidy was set to two because all samples were fertile, thus there were
189 potentially two genetic individuals present (we also conducted the same analyses with the ploidy
190 set to one, without a change in results). Average nucleotide diversity were calculated using
191 VCFtools (Danecek et al. 2011). Linkage distance was calculated between all sites using PLINK

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

192 1.9 and plotted in R with the non-linear least squares smoothing function implemented for the
193 trendline to create a linkage decay chart (Gaunt et al. 2007). Linkage disequilibrium was
194 corrected for using the R package poppr with a threshold of 0.2 and a 1 Kb sliding window
195 (Kamvar et al. 2014). Then, pairwise F_{ST} was calculated among all sampling sites and
196 populations with the linkage disequilibrium corrected dataset using BEDASSLE (Bradburd et al.
197 2015), a Bayesian inference program that estimates the relative influence of ecological and
198 geographic distance on genetic distance. This program automatically excludes sites with missing
199 data when calculating Fst (Bradburd et al. 2015). The dataset was checked for clones by
200 searching for multi-locus genotypes that are >95% identical to account for sequencing and SNP
201 calling errors using the mlg.filter function in the R package popper with a threshold of 0.05
202 (Kamvar et al. 2014).

203 *Statistical Analyses*

204 First, the relationships among populations were explored to determine if there were
205 phylogenetic signals for each distinct sampling site and mountain range, and at what spatial scale
206 the relationships were clearest. Three methods were used to explore population structure:
207 InStruct, Discriminant Analysis of Principle Components (DAPC), and a neighbor-joining tree.
208 InStruct is a Bayesian clustering program that infers self-fertilization rates and clusters
209 individuals into subpopulations using a Markov chain Monte Carlo (MCMC) algorithm (Gao et
210 al. 2007). InStruct was run with K values between one and 10, with five independent chains per
211 K value. Each chain was run with a 100,000 iteration burn-in period, followed by 10,000
212 iterations after burn-in. Mode 2 was used to infer subpopulation structure and selfing rate. The
213 run with the highest DIC was chosen to represent the population structure and inferred selfing
214 rate. We also implemented DAPC, a multivariate approach to identifying genetically distinct

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

215 clusters of individuals, with the clustering algorithm implemented to define groups resulting in
216 10 genetic clusters chosen based on the BIC (Jombart et al. 2010). This method was specifically
217 designed to cope with the large quantity of next generation sequencing data and implemented in
218 R through the package adegenet 2.0 (Jombart 2008). An unrooted neighbor-joining tree was also
219 built to infer the relationships among individuals based on bitwise distances using the R package
220 ape (Paradis et al. 2004).

221 The influence of geographic and ecological distance on genetic distance was investigated
222 using two approaches. First, a partial Mantel test with 10,000 permutations was used to test for
223 correlation between genetic distance, measured as pairwise Fst, and geographic distance
224 (Euclidean distance in kilometers), and a set of four environmental variables. The set of four
225 environmental variables were habitat (boulder in stream vs. exposed rock outcrop) and three non-
226 colinear variables from the widely-used Worldclim dataset: mean temperature of wettest quarter
227 (BIO8), mean temperature of warmest quarter (BIO10), and annual precipitation (BIO12)
228 (Hijmans et al. 2005). These last three variables were retained after removing all correlated
229 climatic variables from Worldclim (see Species Distribution Modeling, below). The habitat
230 variable was based on field observations of the species made while collecting samples. Second, a
231 Bayesian approach as implemented in BEDASSLE was used to estimate the contributions of
232 geographic and ecological distance to genetic distance (Bradburd et al. 2013). The same set of
233 ecological and geographic distance variables were used as input data, along with allele sample
234 sizes and frequencies in all samples. An initial Bayesian analysis, run for 1 million generations,
235 indicated that the effect size of BIO8 and BIO12 were very close to zero, and these were
236 removed from the dataset. A second analysis was run retaining the habitat and BIO10 as
237 environmental variables for 5 million generations with a sample frequency of 10. A third

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

238 analysis was conducted retaining only the habitat as the environmental variable and was run for
239 10 million generations with a sample frequency of 10. Trace plots were examined for
240 convergence, and mean marginal densities and 95% confidence intervals calculated for αE:αD
241 for each environmental variable with the first 50% of generations treated as burn-in and
242 removed.

243 *Reproductive Morphology*

244 To determine whether reproductive structures were observable in *Cetradonia* specimens,
245 fertile samples were dissected to search microscopically for trichogynes, specialized hyphae that
246 receive spermatia (conidia) to begin sexual reproduction, and fertile apothecia, structures that
247 produce fungal spores that are produced through meiosis. Thin sections of five apothecia from
248 three sampling sites were cut by hand with a razor blade through apothecia and mounted on
249 slides. Sections were stained with phloxine and cleared with potassium hydroxide before
250 examination under a compound microscope.

251 *Species Distribution Modeling*

252 Species distribution modeling was used to investigate if the sites with the highest genetic
253 diversity were located in an area that was likely a refugium during the Last Glacial Maximum
254 (LGM). To model past distributions, we first built a species distribution model (SDM) to predict
255 the probability of a species' presence across the landscape for the present, then projected this
256 SDM to past climates. Species distribution modeling was conducted using Maxent v. 3 (Phillips
257 et al. 2006, Phillips and Dudik 2008) after steps were taken to reduce sampling bias and calibrate
258 the model. First, localities were thinned by a 5 km radius to reduce sampling bias by randomly
259 excluding one of two localities when they fell within that radius, as implemented in the R
260 package SpThin (Aiello-Lammens et al. 2015). After thinning, 42 out of the 101 original

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

261 localities were retained and used for all further analyses. The worldclim dataset of 19 bioclimatic
262 variables was used for the environmental data at 10 arc second resolution for the present and 2.5
263 arc minutes for the last glacial maximum. All autocorrelated variables were first removed,
264 leaving mean temperature of wettest quarter (BIO8), mean temperature of warmest quarter
265 (BIO10), and annual precipitation (BIO12) (Hijmans et al. 2005). These three variables were
266 clipped to the extent of the species known range, with a small buffer, for environmental variables
267 from the present and LGM. Two modeling parameters were tuned to identify the best level of
268 complexity for the model: feature classes define the allowed shape of the environmental variable
269 response curves, and the regularization multiplier controls for complexity, with higher values
270 increasingly penalizing complexity (Scheglovitova and Anderson 2013). The best modeling
271 parameters were chosen based on the Akaike Information Criterion corrected for sample size
272 (AICc) (Warren and Seifert 2011). Model tuning was implemented using the R package
273 ENMEval with the 'blocks' setting (Muscarella et al. 2014). The final model was built and
274 projected using all thinned localities with the regularization multiplier set to 3.5 and linear,
275 quadratic, and hinge response curves allowed.

276 **Results**

277 High-coverage, whole-genome shotgun sequencing of one individual of *Cetradonia*
278 *linearis* was obtained and used to assemble a reference genome. Whole-genome shotgun
279 sequencing of 31 additional individuals were mapped to this genome, and the resulting SNPs
280 were used to infer the population structure, biogeographic history, and mating system of the
281 species.

282 *Cetradonia linearis Reference Genome*

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

283 Multiple steps of stringent quality and contaminant filtering resulted in the production of
284 a high-quality, partial reference genome. The original read pool from the sample used to create
285 the reference genome, sample B224, contained 55 million reads, for a total of 16 Gb. The mean
286 PHRED quality score was 33. After trimming, filtering for low quality base calls, and merging
287 paired ends there were 44 million merged reads (where two paired-end reads overlapped and
288 merged into a single sequence) with a total of 5.6 Gb of sequence, and 8.3 million read pairs that
289 did not overlap with a total of 2.1 Gb of sequence. The initial assembly using Minia built 32,669
290 contigs. When contigs under 500 bp were excluded, the total assembly length was 105.5 Mb and
291 the N50 was 3,814 bp. After filtering contaminants, 41.9 million merged reads with 5.4 Gb of
292 sequence remained, as well as 7.6 million paired-end read with 2.0 Gb of sequence. These
293 filtered reads were then assembled, which built 17,199 contigs with a total length of 40.0 Mb and
294 an N50 of 6,093. This assembly was then annotated with protein homology data from *Aspergillus*
295 *niger* ATCC 1015 v4.0, *Cladonia grayi* Cgr/DA2myc/ss v2.0, and *Cochliobolus heterostrophus*
296 C5 v2.0 and ab-initio prediction using SNAP. Then, only contigs for which the annotated gene
297 with the best blastp score against *C. grayi* and *A.niger* proteins most closely matched *C. grayi*
298 were retained for the final reference genome to be used in all downstream analyses. This
299 reference genome was comprised of 2,703 contigs with a total length of 19.5 MB, a contig N50
300 of 10,095 bp, and an average coverage of 54.7 X. CEGMA (Parra et al. 2007) analysis of
301 conserved gene content showed that 74% of universally conserved genes are present in our
302 assembly, suggesting that our assembly is approximately 74% complete. Consistent with this,
303 our assembly was 53-70% as large as the three genomes available for other species in the
304 Cladoniaceae (28 Mb-37 Mb; Armeleo and May 2009; Park et al. 2013). The MAT1-2
305 idiomorph was located in the reference genome, and in 14 of 32 read pools (Supplementary text).

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

306 No MAT1-1 genes were located in any samples. Because only one mating type was discovered
307 we preliminarily determine that the mating system of *C. linearis* may be homothallic, and
308 specifically unisexual (Wilson et al. 2015).

309 *Cetradonia linearis Population Structure*

310 To call SNPs, all read pools were aligned to the reference genome. A total of 126,662
311 SNPs were identified. After correcting for linkage disequilibrium 10,026 SNPs remained. This
312 large reduction in SNPs after correcting for linkage disequilibrium suggests a low rate of
313 recombination. Examination of the linkage decay plot further supports the hypotheses of low
314 recombination rates, as the linkage between sites never falls below 0.2 (Fig. 3). In the dataset
315 used for subsequent analyses the average SNP coverage was 66%, and the coverage per
316 population ranged from 52.6-98.6% (Table 1). Nucleotide diversity (π) within sampling sites
317 ranged from 0.084 for one site in the Great Smoky Mountains, to 0.18 for one site in the Black
318 Mountains. When the samples were grouped by mountain range, nucleotide diversity ranged
319 from 0.148-0.338 (Table 1). Pairwise Fst values between sites ranged from 0.312 to 0.730
320 (Supplementary Table 1).

321 Population structure was first explored through relational analyses. The unrooted NJ tree
322 recovered distinct, mutually exclusive groups that corresponded to distinct mountain ranges (Fig.
323 2). Sampling sites also largely formed mutually exclusive groups, except PV. The one PV
324 sample that did not cluster with other PV samples formed a group with SH, a site that was only
325 1.5 km downstream. The InStruct analysis chain with the highest DIC found seven clusters (Fig.
326 2). Four primary clusters were evident in the results: one that included all samples from the Great
327 Smoky Mountains, one from the southern Nantahalas, one that included the Balsam Mountains
328 and Roan Mountain, and one that included all samples from the Black Mountains (Fig. 2). The

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

329 mean posterior distribution of selfing rates averaged 0.59, with the mean, followed by the
330 variance, for each cluster inferred as follows: 1 = 0.475 (0.085), 2 = 0.511 (0.087), 3 = 0.585
331 (0.074), 4 = 0.606 (0.015), 5=0.620 (0.016), 6 = 0.665 (0.008), 7 = 0.666 (0.014). Ten clusters
332 were found as the most likely grouping of the samples using DAPC. Most clusters were
333 comprised of all individuals from single sampling sites. Group four was the only one that
334 included samples from multiple sites, for a total of 15 individuals from nine sites that included
335 the Great Smoky Mountains, Balsam Mountains, Nantahala Mountains, and Roan Mountain
336 (Supplement Fig. 2). Each of the three sites sampled from the Black Mountains formed their own
337 distinct group, despite their close proximity to each other (1-173 km apart).

338 We tested the influence of geographic versus environmental distance on genetic distance
339 using two methods, and both showed geographic distance as a more significant factor correlating
340 with population structure. First, a partial Mantel Test showed a significant correlation between
341 genetic distance, measured as pairwise Fst, and geographic distance, measured as pairwise
342 Euclidean distance in km, where $r = 0.489$, and $p < 0.01$ (Fig. 4). There were no correlations
343 between genetic distance and any of the environmental distances (Supplementary Table 2). The
344 second analysis was a Bayesian approach implemented in the program BEDASSLE (Bradburd et
345 al. 2013). Here, the relevant value is the ratio of effect size of each environmental variable versus
346 the effect size of the geographic distance (αE:αD). The results were similar to the partial Mantel
347 test, and geographic distance far outweighed the effect of environmental distance. Specifically,
348 the results of the first analysis, which included both habitat and BIO10 as environmental
349 variables, estimated the habitat αE:αD = 0.712 (95% CI = 0.544, 0.978) and the BIO10 αE:αD =
350 0.026 (95% CI = 0.020, 0.036). This can be interpreted as follows: the effect of 10° C mean
351 temperature of the warmest quarter was equal to the effect of 0.026 km of geographic distance,

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

352 and the effect of occurrence in different habitats was equal to 0.75 km of geographic distance. In
353 the second analyses, where only habitat was retained as an environmental variable, the habitat
354 $\alpha E:\alpha D = 0.309$ (95% CI = 0.011, 1.408). Hindcasting the SDM of *Cetradonia linearis* supported
355 the hypothesis that its refugial range was located predominantly in the southern edge of its
356 current range during the LGM (Fig.5). The quality of the SDM was high, with an AUC of 0.919.

357 **Discussion**

358 An understanding of diversification mechanisms in lichens has been hampered by a lack
359 of population genetic studies in these diverse clades of fungi. This study is the first to report the
360 results of a genomic approach for investigating the population structure of a lichen. Low-
361 coverage whole-genome sequencing of lichen fragments produced large quantities of SNP data
362 (>122,000 SNPs) among individuals within a species, even after contaminants were removed by
363 stringent filtering. These results demonstrate that culturing is not required for lichen population
364 genomics. The original hypothesis that the main dispersal strategy of *Cetradonia linearis* is
365 through clonal propagation was not supported, as no clones were identified across or within sites.
366 However, there is evidence that the species less frequently undergoes outcrossing and sexual
367 recombination based on the high rates of linkage disequilibrium (~122K SNPs reduced to ~10K,
368 Fig. 3), estimated selfing rates >0.5, and the putatively unisexual mating system.

369 In contrast to prior work suggesting fungi are not dispersal-limited, the hypothesis that
370 there are low rates of gene flow among populations was supported by high F_{st} values (0.312-
371 0.730), significant correlation between genetic and geographic distance (Mantel Test, $r = 0.489$,
372 $p < 0.01$), and proportionally higher influence of geographic distance on genetic distance when
373 compared to environmental distance ($\alpha E:\alpha D < 1$). There was no evidence for isolation by
374 environment based on the partial Mantel test and BEDASSLE results. However, further studies

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

375 of other environmental variables, such as average high temperatures of warm months or quarters,
376 may reveal signals of adaptation not recovered here. Additionally, future spore trapping and
377 viability assays would be a useful way to directly measure dispersal potential. The sites with the
378 highest genetic diversity were concentrated in the southern portion of the range of the taxon as
379 predicted, suggesting that these may have acted as refugial areas during the LGM. The results of
380 this study support the notion that gene flow among fungal populations decreases with distance -
381 populations show strong signs of IBD despite the production of small propagules (O'Malley
382 2007) - suggesting that genetic drift may represent a more important process in diversification of
383 lichenized fungi than previously appreciated. Our data also indicate that recombination can be
384 low despite the frequent presence of sexual spore producing structures. This finding highlights
385 the phenomenon that observed reproductive mode does not necessarily translate directly to the
386 frequency of recombination (Taylor et al. 2015).

387 *Influence of Reproductive Strategy on Population Genetic Structure in Lichens*

388 Our data on *Cetradonia linearis* contribute to a growing understanding of the relationship
389 between fungal reproductive types and genetic diversity and structure. Three species of
390 lichenized fungi were previously examined with detailed population genetic studies including
391 analysis of the mating-system (Itten and Honegger 2010; Singh et al. 2012; Alors et al. 2017).
392 One of these species was investigated with RAPD-PCR and the other two with microsatellites, so
393 comparisons among the studies, and with our study based on genomic data, must be done with
394 consideration of the very different underlying data. Nonetheless, because there are no genomics
395 studies on lichen population genetic structure currently published, a careful comparison of our
396 results with this previous research is useful. The lichen *Xanthoria parietina* was found to be
397 unisexual, having only the MAT1-2 gene present in all individuals investigated, and no observed

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

398 instances of trichogynes, though it is almost always fertile (Scherrer et al. 2005). The population
399 genetic structure of *X. parietina* based on RAPD-PCR fingerprinting revealed high rates of
400 genotypic diversity within populations, even on a microsites scale, and much lower genetic
401 diversity between populations than within them (Itten and Honegger 2010). A study of the lichen
402 *Parmelina carporrhizans* found a similar pattern of very high gene flow among most populations
403 sampled, though it is a heterothallic species (Alors et al. 2017). The pattern observed in these
404 two species starkly contrasts with that of *Lobaria pulmonaria*, a heterothallic species that is often
405 observed without sexual reproductive structures, in which apothecia usually are not produced
406 until individuals are 15-25 years old (Denison 2003; Høistad and Gjerde 2011; Singh et al.
407 2012). *Lobaria pulmonaria* consistently displays high rates of clonality within populations and
408 sampling sites (Werth et al. 2006; Sing et al. 2012). One way to explain the difference between
409 the population genetic patterns of the two heterothallic species is the ratio of the two alternate
410 MAT idiomorphs: *L. pulmonaria* ratios are often skewed in populations while *P. carporrhizans*
411 populations have equal ratios (Singh et al. 2012; Alors et al. 2017). Population genetic structure
412 and biology of *Cetradonia linearis* is more similar to *X. parietina* and *P. carporrhizans* because
413 1) it is almost always fertile, 2) no clones have been identified (defined as >95% shared SNPs for
414 the genomic data), even from closely collected colonies, and 3) there is a high level of
415 polymorphism within each population ($\pi = 0.148 – 0.338$). However, our results indicate *C.*
416 *linearis* populations have low rates of gene flow, which contrasts with the pattern of little genetic
417 structure found in both *X. parietina* and *P. carporrhizans*. Additional studies using genomics to
418 investigate population genetics in lichenized fungi will allow more direct comparisons of genetic
419 structure across diverse fungal clades. These results, along with the high rate of linkage

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

420 disequilibrium, suggest that while *C. linearis* does not seem to frequently reproduce clonally,
421 there must be some rate of self-fertilization or clonality and dispersal restriction that leads to the
422 genetic isolation of populations. To draw large-scale conclusions about the influence of observed
423 reproductive mode on recombination, further studies tackling a greater breadth of taxonomic
424 sampling throughout lichenized fungi will be required. Specific efforts are needed to target
425 phylogenetic, morphological, ecological, and reproductive diversity to determine what factors
426 most strongly shape recombination.

427 *Biogeographic History*

428 The southern Appalachian Mountain Range is one of the oldest continuously exposed
429 land masses on earth, and has served as a refugium at multiple points in geological history
430 (Braun 1950). Thus, though it is a relatively small area, the long and complex geological history
431 of the region has shaped similarly strong, complex population genetics patterns in endemic
432 species across multiple domains of life (Manos and Meireles 2015). The population genetics of
433 *C. linearis* are no exception. The southern portion of the current extent of *C. linearis* was likely a
434 refugium during Pleistocene glaciation. The lines of evidence to support this hypothesis include
435 the observation that genetic diversity is higher in southern populations and location of suitable
436 areas predicted by the hindcast SDM (Fig. 5). Interestingly, the model also suggested an
437 expansion of the range to lower elevation areas (Fig. 5). This finding is consistent with
438 hypotheses that ranges of present-day high-elevation endemics expanded downslope during
439 Pleistocene glaciation (Crespi et al. 2003; Bruhl 1997; Premoli et al. 2007; Desamore et al.
440 2010). While this downslope expansion might have been expected to connect populations and
441 diminish the signal of IBD, the data generated for this study still show a strong geographic signal
442 of increasing structure with distance. Population genetic studies of other high-elevation, southern

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

443 Appalachian endemics showed similarly strong signals of IBD, including the salamander
444 *Desmognathus wrightii* (Crespi et al. 2003) and the spider *Hypochilus pococki* (Keith and Hedin
445 2012). A further parallel between the genetic structures of *D. wrightii* and *C. linearis* is that
446 Roan Mountain populations did not group with populations from the Black Mountains, despite
447 their close geographic proximity (Crespi et al. 2003). Population differentiation was so strong in
448 *Hypochilus pococki* that the authors suggested it may actually be comprised of multiple cryptic
449 species (Keith and Hedin 2012). The long and complex geological history of the southern
450 Appalachians has resulted in not only high levels of species diversity, but also high genetic
451 diversity within species.

452 **Conclusion**

453 The results presented here provide strong evidence that the rare, narrowly endemic
454 fungus *Cetradonia linearis* has highly geographically isolated populations over the small area of
455 its distribution (Fig. 2). The population structure of *C. linearis* is congruent with other high-
456 elevation southern Appalachian endemics, suggesting that dispersal among mountain peaks in
457 the region is not frequent for multiple groups of organisms (Crespi et al. 2003; Keith and Hedin
458 2012; Fig. 5). We found no evidence of clones in our sampling, however we did find evidence
459 for low rates of recombination, possibly facilitated by a homothallic reproductive system that
460 allows self-fertility (Fig. 3). These results support a growing body of literature suggesting that
461 fungal dispersal can be limited across relatively small spatial scales, despite the production of
462 very small propagules (O'Malley 2007; Taylor et al. 2012). While other studies have found a
463 strong influence for environmental factors influencing population structure of fungi (Branco et
464 al. 2015), we found no evidence for isolation by environment (Supplement Table 1). Future
465 comparative studies are required to full understand how extrinsic and intrinsic factors shape the

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

466 population structure and recombination rates of fungi with different ecological requirements, life
467 histories, and reproductive strategies (Grünwald et al. 2016; Peter and Schacherer 2016). These
468 studies will be facilitated by rapid advancements in population genomics methods, which
469 promise to reshape current perspective on fungal biology.

470 **Acknowledgements**

471 We would like to thank the following people for aiding with this research: Alex Cecil and Jenna
472 Dorey for their field assistance, Gary Kauffman for help locating populations throughout the
473 National Forest and many helpful discussions about *Gymnoderma*, Chris Ulrey for help locating
474 the species on Blue Ridge National Park land, and especially for rapelling expertise, Dr. Richard
475 Harris for conducting microscopy to search for trichogynes and spores. Library preparations and
476 sequencing was conducted at The Rockefeller Genome Resource Center. Funding for this
477 research came from Highlands Biological Station, NSF GRFP, and NSF DEB#1145511.

478 **Literature Cited**

479 Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., Anderson, R. P. 2015.
480 spThin: an R package for spatial thinning of species occurrence records for use in
481 ecological niche models. *Ecography* 38: 541–545.

482 Allen, J. L., Lendemer, J.C., McMullin, T. 2015. Cetradonia linearis. *The IUCN Red List of*
483 *Threatened Species 2015* e.T70386009A70386019.

484 Alors, D., Grande, F. D., Cubas, P. *et al.* 2017. Panmixia and dispersal from the Mediterranean
485 Basin to Macaronesian Islands of a macrolichen species. *Sci. Rep* **7**: 40879.

486 Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. 1990. Basic local alignment
487 search tool. *J. Mol. Biol.* 215: 403–410.

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

488 Andersen, M. R., Salazar, M. P., Schaap, P. J. *et al.* 2011. Comparative genomics of citric-acid-
489 producing *Aspergillus niger* ATCC 1015 versus enzyme-producing CBS 513.88.*Genome*
490 *Res.* 21: 885–897.
491 Armaleo, D., May, S. 2009. Sizing the fungal and algal genomes of the lichen *Cladonia*
492 *grayi* through quantitative PCR. *Symbiosis* 49: 43.
493 Bolger, A. M., Lohse, M., Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina
494 sequence data. *Bioinformatics* 30: 2114–2120.
495 Bradburd, G. S., Ralph, P. L., Coop, G. M. 2013. Disentangling the effects of geographic and
496 ecological isolation on genetic differentiation. *Evolution* 67: 3258–3273.
497 Branco, S., Gladieux, P., Ellison, C. E. *et al.* 2015. Genetic isolation between two recently
498 diverged populations of a symbiotic fungus. *Mol Ecol* 24: 2747–2758.
499 Braun, L. E. 1950. *Deciduous Forests of Eastern North America*. The Blackburn Press, Caldwell,
500 NJ.
501 Brodo, I. M., Sharnoff, M. S. D., Sharnoff, S. 2001. *Lichens of North America*. Yale University
502 Press, New Haven, CT.
503 Bruhl, C. A. 1997. Flightless insects: A test case for historical relationships of African
504 mountains. *J. Biogeogr* 24: 233–250.
505 Cantarel, B. L., Korf, I., Robb, S. M. C. *et al.* 2008. MAKER: an easy-to-use annotation pipeline
506 designed for emerging model organism genomes. *Genome Res.* 18: 188–196.
507 Chikhi, R., Rizk, G. 2013. Space-efficient and exact de Bruijn graph representation based on a
508 Bloom filter. *Algorithms Mol. Biol.* 8: 22.

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

509 Condon, B. J., Leng, Y., Wu, D. *et al.* 2013. Comparative genome structure, secondary
510 metabolite, and effector coding capacity across *Cochliobolus* pathogens. *PLoS Genet.* 9:
511 e1003233.
512 Crespi, E. J., Rissler, L. J., Browne, R. A. 2003. Testing Pleistocene refugia theory:
513 phylogeographical analysis of *Desmognathus wrighti*, a high-elevation salamander in the
514 southern Appalachians. *Mol. Ecol.* 12: 969–984.
515 Crittenden, P. D., David, J. C., Hawksworth, D. L., Campbell, F. S. 1995. Attempted isolation
516 and success in the culturing of a broad spectrum of lichen-forming and lichenicolous
517 fungi. *New Phytol.* 130: 267–297.
518 Dal Grande, F., Widmer, I., Beck, A., Scheidegger, C. 2010. Microsatellite markers
519 for *Dictyochloropsis reticulata* (Trebouxiophyceae), the symbiotic alga of the
520 lichen *Lobaria pulmonaria* (L.). *Conserv. Genet.* **11:** 1147–1149.
521 Danecek, P., Auton, A., Abecasis, G. *et al.* 2011. The variant call format and
522 VCFtools. *Bioinformatics* 27: 2156–2158.
523 Denison, W. C. 2003. Apothecia and ascospores
524 of *Lobaria oregana* and *Lobaria pulmonaria* investigated. *Mycologia* 95: 513–518.
525 Desamore, A., Vanderpoorten, A., Laenen, B., Gradstein, S. R., Kok, P. J. R. 2010.
526 Biogeography of the Lost World (Pantepui region, northeastern South America): Insights
527 from bryophytes. *Phytotaxa* 9: 254–265.
528 Devkota, S., Cornejo, C., Werth, S., Chaudhary, R. P., Scheidegger, C. 2014. Characterization of
529 microsatellite loci in the Himalayan lichen fungus *Lobaria*
530 *pindarensis*(Lobariaceae). *Appl. Plant Sci.* 2: 1300101.

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

531 Dyer, P. S. 2008. Evolutionary biology: genomic clues to original sex in fungi. *Curr. Biol.* 18:
532 R207–9.
533 Gao, H., Williamson, S., Bustamante, C. D. 2007. A Markov chain Monte Carlo approach for
534 joint inference of population structure and inbreeding rates from multilocus genotype
535 data. *Genetics* 176: 1635–1651.
536 Garrison, E., Marth, G. 2012. Haplotype-based variant detection from short-read
537 sequencing. *arXiv [q-bio.GN]*.
538 Gladieux, P., Ropars, J., Badouin, H. *et al.* 2014. Fungal evolutionary genomics provides insight
539 into the mechanisms of adaptive divergence in eukaryotes. *Mol. Ecol.* 23: 753–773.
540 Gaunt, T. R., Rodríguez, S., Day, I. N. 2007. Cubic exact solutions for the estimation of pairwise
541 haplotype frequencies: implications for linkage disequilibrium analyses and a web tool
542 "CubeX." *BMC Bioinformatics* 8: 428.
543 Grünwald, N. J., McDonald, B. A., Milgroom, M. G. 2016. Population Genomics of Fungal and
544 Oomycete Pathogens. *Annu. Rev. Phytopathol*. 54: 323–346.
545 Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., Jarvis, A. 2005. Very high resolution
546 interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25: 1965–1978.
547 Høistad, F., Gjerde, I. 2011. *Lobaria pulmonaria* can produce mature ascospores at an age of less
548 than 15 years. *Lichenologist* 43: 495–497.
549 Itten, B., Honegger, R. 2010. Population genetics in the homothallic lichen-forming
550 ascomycete *Xanthoria parietina*. *Lichenologist* 42: 751–761.
551 Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic
552 markers. *Bioinformatics* 24: 1403–1405.

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

553 Jombart, T., Devillard, S., Balloux, F. 2010. Discriminant analysis of principal components: a
554 new method for the analysis of genetically structured populations. *BMC Genetics*11: 94.
555 Kamvar, Z. N., Tabima, J. F., Grünwald, N. J. 2014. Poppr: an R package for genetic analysis of
556 populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2: e281.
557 Keith, R., Hedin, M. 2012. Extreme mitochondrial population subdivision in southern
558 Appalachian paleoendemic spiders (Araneae: Hypochilidae: *Hypochilus*), with
559 implications for species delimitation. *J. Arachnol.* 40: 167–181.
560 Korf, I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59
561 Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M., Blaxter, M. 2013. Blobology: exploring raw
562 genome data for contaminants, symbionts and parasites using taxon-annotated GC-
563 coverage plots. *Front. Genet.* 4: 237.
564 Leducq, J.-B., Nielly-Thibault, L., Charron, G. *et al.* 2016. Speciation driven by hybridization
565 and chromosomal plasticity in a wild yeast. *Nat. Microbiol.* 1: 15003.
566 Lee, S. C., Ni, M., Li, W., Shertz, C., Heitman, J. 2010. The evolution of sex: a perspective from
567 the fungal kingdom. *Microbiol. Mol. Biol. Rev.* 74: 298–340.
568 Li, H., Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler
569 transform. *Bioinformatics* 25: 1754–1760.
570 Linde, C. C., Zala, M., Ceccarelli, S., McDonald, B. A. 2003. Further evidence for sexual
571 reproduction in *Rhynchosporium secalis* based on distribution and frequency of mating-
572 type alleles. *Fungal Genet. Biol.* 40: 115–125.
573 Lindgren, H., Leavitt, S. D., Lumbsch, T. 2016. Characterization of microsatellite markers in the
574 cosmopolitan lichen-forming *fungus Rhizoplaca*
575 *melanophthalma*(Lecanoraceae). *MycoKeys*, 14: 31–36.

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

576 Liti, G., Carter, D. M., Moses, A. M. *et al.* (2009) Population genomics of domestic and wild
577 yeasts. *Nature* 458: 337–341.
578 Lücking, R., Hodkinson, B. P., Leavitt, S. D. 2016. The 2016 classification of lichenized fungi in
579 the Ascomycota and Basidiomycota – Approaching one thousand genera.*Bryologist* 119:
580 361–416.
581 Lutsak, T., Fernández-Mendoza, F., Greshake, B. *et al.* 2016. Characterization of microsatellite
582 loci in the lichen-forming fungus *Cetraria aculeata* (Parmeliaceae, Ascomycota).*Appl.*
583 *Plant Sci.* 4: 1600047.
584 Magain, N., Forrest, L. L., Sérusiaux, E., Goffinet, B. 2010. Microsatellite primers in
585 the *Peltigera dolichorhiza* complex (lichenized ascomycete, Peltigerales). *American*
586 *Journal of Botany*, **97**, e102–4.
587 Magoč, T., Salzberg, S. L. 2011. FLASH: fast length adjustment of short reads to improve
588 genome assemblies. *Bioinformatics* 27: 2957–2963.
589 Manos, P. S., Meireles, J. E. 2015. Biogeographic analysis of the woody plants of the Southern
590 Appalachians: Implications for the origins of a regional flora. *Am. J. Bot.* 102: 780–804.
591 Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing
592 reads. *EMBnet.journal* 17: 10–12.
593 McDonald, T. R., Mueller, O., Dietrich, F. S., Lutzoni, F. 2013. High-throughput genome
594 sequencing of lichenizing fungi to assess gene loss in the ammonium
595 transporter/ammonia permease gene family. *BMC Genomics* 14: 225.
596 Muscarella, R., Galante, P. J., Soley-Guardia, M. *et al.* 2014. ENMeval: An R package for
597 conducting spatially independent evaluations and estimating optimal model complexity
598 for Maxent ecological niche models. *Methods Ecol. Evol.* 5: 1198–1205.

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

599 Nadyeina, O., Dymytrova, L., Naumovych, A. *et al.* 2014. Microclimatic differentiation of gene
600 pools in the *Lobaria pulmonaria* symbiosis in a primeval forest landscape. *Mol. Ecol.* 23:
601 5164–5178.
602 Nadyeina, O., Cornejo, C., Boluda, C. G. *et al.* 2014. Characterization of microsatellite loci in
603 lichen-forming fungi of *Bryoria* section *Implexae* (Parmeliaceae). *Appl. Plant Sci.*2:
604 1400037.
605 O'Malley, M. A. 2007. The nineteenth century roots of "everything is everywhere." *Nat. Rev.*
606 *Microbiol.* 5: 647–651.
607 Ohm, R. A., Feau, N., Henrissat, B. *et al.* 2012. Diverse lifestyles and strategies of plant
608 pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS*
609 *Pathog.*8: e1003037.
610 Paoletti, M., Seymour, F. A., Alcocer, M. J. C. *et al.* 2007. Mating type and the genetic basis of
611 self-fertility in the model fungus *Aspergillus nidulans*. *Curr. Biol.* 17: 1384–1389.
612 Papadopoulos, J. S., Agarwala, R. 2007. COBALT: constraint-based alignment tool for multiple
613 protein sequences. *Bioinformatics* 23. 1073–1079.
614 Paradis, E., Claude, J., Strimmer, K. 2004. APE: Analyses of Phylogenetics and Evolution in R
615 language. *Bioinformatics* 20: 289–290.
616 Park, S.-Y., Choi, J., Kim, J. A. *et al.* 2013. Draft Genome Sequence of *Cladonia*
617 *macilenta* KoLRI003786, a Lichen-Forming Fungus Producing Biruloquinone. *Genome*
618 *Announc.* 1: e00695-13
619 Parra, G., Bradnam, K., Korf, I. 2007. CEGMA: a pipeline to accurately annotate core genes in
620 eukaryotic genomes. *Bioinformatics* 23: 1061–1067.

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

621 Peter, J., Schacherer, J. 2016. Population genomics of yeasts: towards a comprehensive view
622 across a broad evolutionary scale. *Yeast* 33: 73–81.
623 Phillips, S. J., Dudík, M., Schapire, R. E. 2004. A Maximum Entropy Approach to Species
624 Distribution Modeling. *Proceedings of the Twenty-First International Conference on*
625 *Machine Learning*. 655–662.
626 Phillips, S. J., Anderson, R. P., Schapire, R. E. 2006. Maximum entropy modeling of species
627 geographic distributions. *Ecol. Modell.* 190: 231–259.
628 Premoli, A. C., del Castillo, R. F., Newton, A. C. *et al.* 2007. Patterns of Genetic Variation in
629 Tree Species and their Implications for Conservation. In: *Biodiversity Loss and*
630 *Conservation in Fragmented Forest Landscapes: the Forests of Montane Mexico and*
631 *Temperate South America*. AC Newton, Ed. CABI, Wallingford, UK.
632 Scherrer, S., Zippler, U., Honegger, R. 2005. Characterisation of the mating-type locus in the
633 genus *Xanthoria* (lichen-forming ascomycetes, Lecanoromycetes). *Fungal Genet.*
634 *Biol.* 42: 976–988.
635 Schoch, C. L., Sung, G.-H., López-Giráldez, F. *et al.* 2009. The Ascomycota tree of life: a
636 phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive
637 and ecological traits. *Syst. Biol.* 58: 224–239.
638 Scheglovitova, M., Anderson, R. P. 2013. Estimating optimal complexity for ecological niche
639 models: A jackknife approach for species with small sample sizes. *Ecol. Modell.*269: 9–
640 17.
641 Seymour, F. A., Crittenden, P. D., Dickinson, M. J. *et al.* 2005. Breeding systems in the lichen-
642 forming fungal genus Cladonia. *Fungal Genet. Biol.* 42: 554–563.

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

643 Simpson, J. T., Wong, K., Jackman, S. D. *et al.* 2009 ABySS: a parallel assembler for short read
644 sequence data. *Genome Res.* 19: 1117–1123.
645 Singh, G., Dal Grande, F., Cornejo, C., Schmitt, I., Scheidegger, C. 2012. Genetic basis of self-
646 incompatibility in the lichen-forming fungus *Lobaria pulmonaria* and skewed frequency
647 distribution of mating-type idiomorphs: implications for conservation. *PloS one* **7**:
648 e51402.
649 Singh, G., Ashby, A. M. 1998. Cloning of the mating type loci
650 from *Pyrenopeziza brassicae* reveals the presence of a novel mating type gene within a
651 discomycete MAT 1-2 locus encoding a putative metallothionein-like protein. *Mol.*
652 *Microbiol.* 30: 799–806.
653 Spielman, D., Brook, B. W., Briscoe, D. A., Frankham, R. 2004. Does inbreeding and loss of
654 genetic diversity decrease disease resistance? *Conserv. Genet.* 5: 439–448.
655 Stefanini, I., Dapporto, L., Berná, L. *et al.* 2016. Social wasps are a *Saccharomyces* mating
656 nest. *Proc. Natl. Acad. Sci. USA* 113: 2247–2251.
657 Taylor, J. W., Hann-Soden, C., Branco, S., Sylvain, I., Ellison, C. E. 2015. Clonal reproduction
658 in fungi. *Proc. Natl. Acad. Sci. USA* 112: 8901–8908.
659 Tsai, I. J., Bensasson, D., Burt, A., Koufopanou, V. 2008. Population genomics of the wild
660 yeast *Saccharomyces paradoxus*: Quantifying the life cycle. *Proc. Natl. Acad. Sci.*
661 *USA*105: 4957–4962.
662 United States Fish and Wildlife Service (USFWS) (2013) *Rock Gnome Lichen (Gymnoderma*
663 *lineare) 5-year Review: Summary and Evaluation*. USFWS.

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

664 Waalwijk, C., Mendes, O., Verstappen, E. C. P., de Waard, M. A., Kema, G. H. J. 2002.
665 Isolation and characterization of the mating-type idiomorphs from the wheat septoria leaf
666 blotch fungus *Mycosphaerella graminicola*. *Fungal Genet Biol.* 35: 277–286.
667 Warren, D. L., Seifert, S. N. 2011. Ecological niche modeling in Maxent: the importance of
668 model complexity and the performance of model selection criteria. *Ecol. Appl.* 21: 335–
669 342.
670 Wei, J., Ahti, T. 2002. *Cetradonia*, a New Genus in the New Family Cetradoniaceae
671 (Lecanorales, Ascomycota). *Lichenologist* 34: 19–31.
672 Werth, S., Wagner, H. H., Holderegger, R., Kalwij, J. M., Scheidegger, C. 2006. Effect of
673 disturbances on the genetic diversity of an old-forest associated lichen. *Molecular*
674 *Ecology*, **15**, 911–921.
675 Widmer, I., Dal Grande, F., Cornejo, C., Scheidegger, C. 2010. Highly variable microsatellite
676 markers for the fungal and algal symbionts of the lichen *Lobaria pulmonaria* and
677 challenges in developing biont-specific molecular markers for fungal
678 associations. *Fungal Biol.* 114: 538–544.
679 Wilson, A. M., Wilken, P. M., van der Nest, M. A. *et al.* 2015. Homothallism: an umbrella term
680 for describing diverse sexual behaviours. *IMA fungus* 6: 207–214.
681 Wyss, T., Masclaux, F. G., Rosikiewicz, P., Pagni, M., Sanders, I. R. 2016. Population genomics
682 reveals that within-fungus polymorphism is common and maintained in populations of
683 the mycorrhizal fungus *Rhizophagus irregularis*. *ISME J.* 10: 2514–2526.
684 Zhou, Q.-M., Wei, J.-C., Ahti, T., Stenroos, S., Högnabba, F. 2006. The systematic position
685 of *Gymnoderma* and *Cetradonia* based on SSU rDNA sequences. *J. Hattori Bot.*
686 *Lab.*100: 871–880.

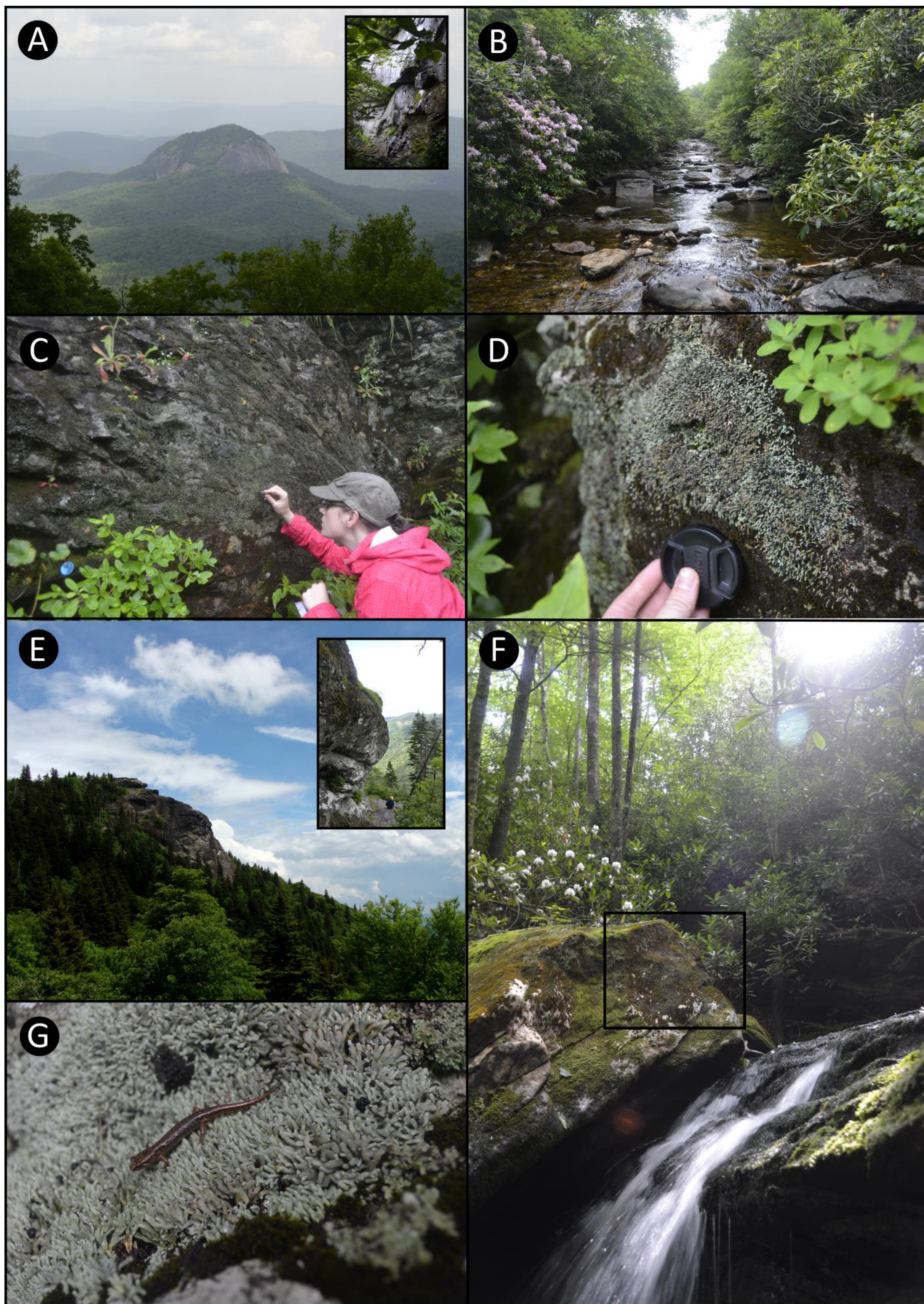
bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

687 **Table 1.** Site names, mountain range, individuals sampled, average percent SNPs covered for
688 each site/mountain range (% Cov), and average nucleotide diversity for all sampled sites and
689 mountain ranges.

Mtn. Range	# Ind.	% Cov	π
Smokies	5	58.1	0.338
Nantahalas	12	52.6	0.329
Blacks	6	98.6	0.268
Balsams	4	71.9	0.208
Roan	3	64.4	0.148
All	32	66	0.328

690

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.



691

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

Figure 1. Morphology, habit, and habitat of *Cetradonia linearis*. A) Large granite dome where species occurs at base of large rock faces, inset shows seeping rock faces where the species occurs; B) Stream habitat where species occurs frequently on scattered rocks and boulders throughout; C) Large boulder face covered in the species illustrating sampling protocol using sterile forceps; D) Fertile colony on mossy boulder in stream; E) Large rock outcrop hosting colonies of the species, inset shows view from *Cetradonia linearis* perspective; F) Waterfall populations are very abundant, one colony outlined by black box; G) Colony displaying apothecium and potential zoochory event.

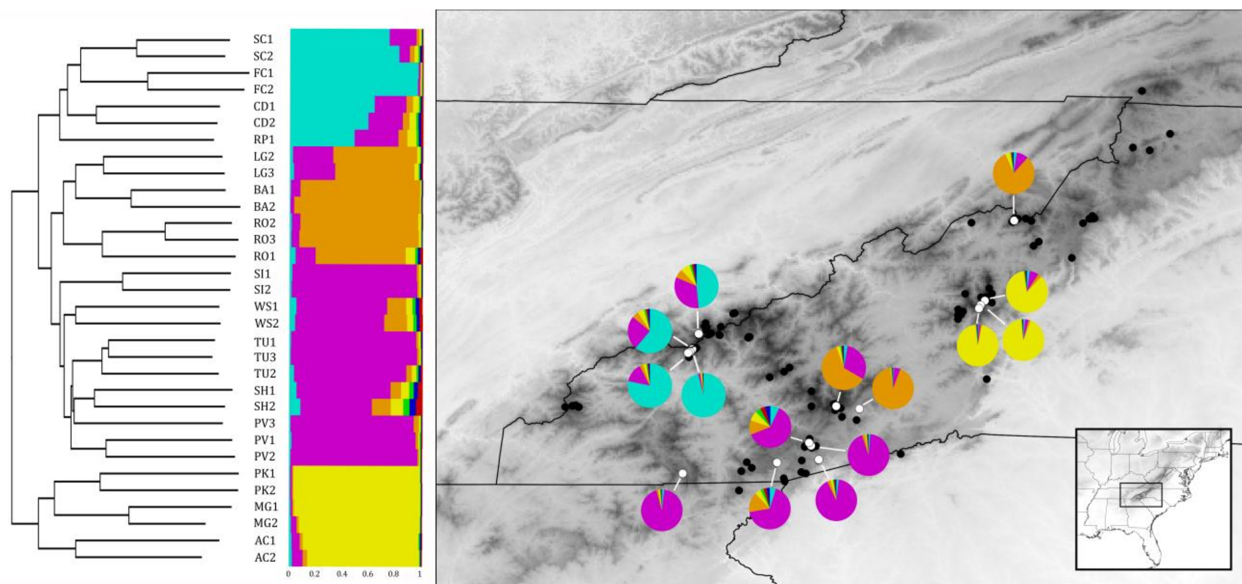
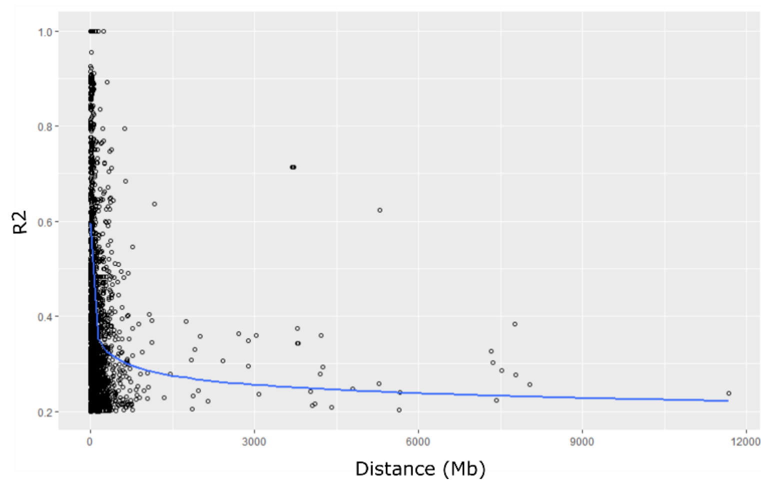
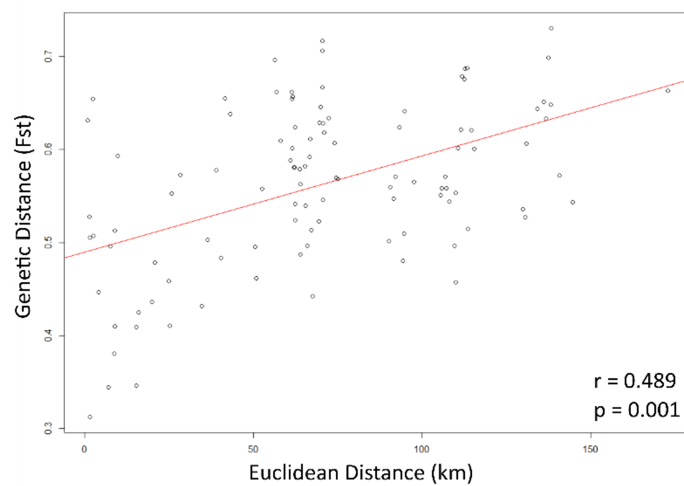


Figure 2. Population structure of *Cetradonia linearis*. Left: Neighbor-joining tree showing hierarchical clustering of all sampled individuals. Middle: Proportional cluster belonging of each individual sampled as inferred by InStruct. Right: Distribution of *C. linearis* and average proportional cluster belonging for each sampling site.

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

706 

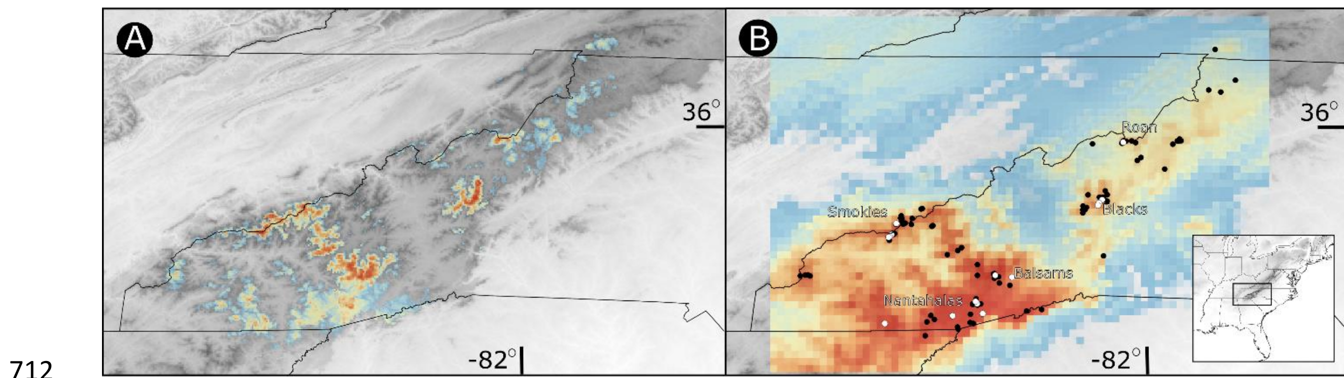
707 **Figure 3.** Linkage decay plot for *Cetradonia linearis*.

708 

709 **Figure 4.** Scatterplot of genetic vs. geographic distance and outcome of statically significant
710 partial Mantel test.

711

bioRxiv preprint doi: https://doi.org/10.1101/237164; this version posted December 21, 2017. 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.

712 

713 **Figure 5.** Species distribution model of *Cetradonia linearis* A) in the present; B) at the last
714 glacial maximum. Probability of *C. linearis* grades from blue (low) to red (high). Inset gives
715 larger spatial orientation of study area.

716