bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1

4

5

6

7

8 **Pronounced Genetic Separation Among Varieties of the *Primula cusickiana* Species**

9 **Complex, a Great Basin Endemic**

10

11

12 **Austin Koontz,[1,4] William D. Pearse,[2] and Paul Wolf [3]**

13

14 [1]Department of Biology, Utah State University, Logan, Utah, 84322, USA;

15 austin.koontz@usu.edu

16

17 [2]Department of Life Sciences, Imperial College London, Silwood Park Campus,

18 Buckhurst Rd., Ascot, Berkshire SL5 7PY, UK

19

20 [3]Department of Biological Sciences, University of Alabama in Huntsville, Huntsville,

21 Alabama, 35899, USA

22

23 [4]Author for correspondence

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

24 ***Abstract***—Distinguishing between unique species and populations with
25 strong genetic structure is a common challenge in population genetics, especially in
26 fragmented habitats where allopatric speciation may be widespread and distinct
27 groups may be morphologically similar. Such is often the case with species
28 complexes across sky island environments. In these scenarios, biogeography may
29 help to explain the relations between species complex members, and RADseq
30 methods are commonly used to compare closely related species across thousands of
31 genetic loci. Here we use RADseq to clarify the relations between geographically
32 distinct but morphologically similar varieties of the *Primula cusickiana* species
33 complex, and to contextualize past findings of strong genetic structure among
34 populations within varieties. Our genomic analyses demonstrate pronounced
35 separation between isolated populations of this Great Basin endemic, indicating that
36 the current varietal classification of complex members is inaccurate and
37 emphasizing their conservation importance. We discuss how these results
38 correspond to recent biogeographical models used to describe the distribution of
39 other sky island taxa in western North America. Our findings also fit into a wider
40 trend observed for alpine *Primula* species complexes, and we consider how
41 heterostylous breeding systems may be contributing to frequent diversification via
42 allopatric speciation in this genus.

43 ***Keywords***— allopatry, biogeography, cryptic speciation, Great Basin,
44 heterostyly, populations, *Primula*, RADseq, sky island

45

46

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

47 A canonical driver of biological diversification is allopatry, whereby geographic
48 barriers lead to population isolation and, eventually, speciation. Sky islands are places
49 where sharp changes in elevation lead to pronounced ecological differences over
50 relatively short distances, providing the types of barriers required for allopatric speciation
51 to take place. Historically, climatic fluctuations have determined the presence and
52 distribution of sky island environments for mountain ranges across the world, and this in
53 turn is reflected by the genetic patterns seen in montane species today (Hewitt 2000).
54 However, in this biogeographic context, distinguishing between closely related species
55 and genetically structured populations may prove challenging (Huang 2020), especially if
56 similar niches across mountain ranges maintain phenotypic similarities (e.g. Yang et al.
57 2019). Additionally, in the short-term, genetic patterns will be influenced by particular
58 aspects of a species' biology, such as dispersal and breeding systems, which may
59 facilitate or hinder reproductive isolation between genetically distinct entities. Here, we
60 examine the genomic relations between the sky island populations of members of the
61 *Primula cusickiana* species complex, a group of plants endemic to the Great Basin region
62 of the western United States.

63 The *P. cusickiana* species complex is a group of herbaceous, perennial plants that
64 fall within the Parryi section of *Primula*. The morphological differences between the four
65 complex varieties—*maguirei*, *cusickiana*, *nevadensis*, and *domensis* (see Fig. 1)—are
66 subtle: *maguirei* (Williams 1936) and *cusickiana* (Gray 1888) are entirely glabrous, and
67 distinguished from one another by relative calyx length, while in *nevadensis* (Holmgren
68 1967) and *domensis* (Kass and Welsh 1985), plants are pubescent and have slightly
69 different corolla tube lengths (Holmgren and Kelso 2001; Holmgren et al. 2005). Despite

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

70 these subtle differences, varieties *cusickiana*, *nevadensis*, and *maguirei* were originally
71 classified as separate species, based on ecological traits and distinct geographic ranges.
72 The discovery and publication of *P. domensis* in 1985, along with the continued
73 collection of the other varieties, began to cast doubt on the species distinction for each
74 complex member. A 2001 review determined that the morphological differences were
75 insufficient for species classification, and subsumed each species to the level of variety
76 (Holmgren and Kelso 2001).

77 At the time of this shift, no genetic data was available to justify classification at
78 the variety level. However, a 1997 analysis of variety *maguirei* used allozyme marker
79 genes to uncover a significant degree of genetic structure between the relatively
80 proximate (~10 km) populations (Wolf and Sinclair 1997) within this one taxon. A later
81 analysis of the same populations using amplified fragment length polymorphism (AFLP)
82 loci confirmed this finding, and found similar levels of polymorphism between the upper
83 and lower canyon groups, suggesting this genetic structure is not the result of a past
84 bottleneck event (Bjerregaard and Wolf 2004). A further analysis of AFLP and
85 chloroplast DNA from the *Primula* section Parryi showed *maguirei* and the other *P.*
86 *cusickiana* complex members as being monophyletic, but relationships within the
87 complex were incongruent, with only weak support of a clade containing *nevadensis* and
88 *domensis* being sister to a clade made up of *maguirei* and *cusickiana* (Kelso et al. 2009).
89 To better resolve the relationships between varieties, the authors suggested an analysis
90 utilizing more populations from across the range of this species complex. Restriction-site
91 associated sequencing (RADseq) technologies available today, with their ability to
92 generate reads over many sequence regions of closely related individuals, are well-suited

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

93 to provide the data required for such an analysis.

94 In addition to clarifying the genetic relations between geographically distinct
95 varieties, a more detailed analysis of the *P. cusickiana* species complex can meaningfully
96 contribute to ongoing conservation efforts. Variety *maguirei* was listed as Threatened in
97 1985, due to its unique habitat in Logan Canyon and threats of habitat loss due to
98 development (Fish and Wildlife Service 1985). Given the strong genetic structure
99 between *maguirei*'s populations, either population may be more closely related to
100 populations of a different complex variety than the neighboring Logan Canyon
101 population—a finding which would have significant implications for the protection of
102 this variety. More broadly, an understanding of the genomic relations at the species
103 complex level will determine whether the varietal classification properly reflects the
104 extent of genomic divergence of each complex member, and thus the extent of unique
105 evolutionary history. This understanding can direct management of the narrow-range
106 endemics included in this species complex—such as *maguirei*, but also *nevadensis* and
107 *domensis*—and also inform the identification of potential evolutionary significant units
108 (Coates et al. 2018).

109 We sought to clarify the relatedness of *P. cusickiana* complex members by using
110 a RADseq approach to genotype all four varieties located at distinct populations scattered
111 throughout the Great Basin. In addition to contextualizing the genetic structure between
112 the upper and lower Logan Canyon *maguirei* populations, this analysis provides insights
113 into the biogeographic history of this species complex, and could have important
114 conservation implications for this rare endemic plant.

115

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

116 ## MATERIALS AND METHODS

117 ***Sampling***—All *P. cusickiana* species complex samples were gathered in the
118 field, along with samples of *P. parryi* (Gray 1888), which was used as an outgroup in
119 genetic analyses. Populations and their respective flowering times were determined
120 using herbarium specimens, and collection sites were selected to maximize the
121 geographic distribution of each variety. At each population location, an individual
122 plant was removed as completely as possible as a voucher specimen. For DNA
123 samples, two leaves from each of ten plants were removed and placed in labeled
124 paper envelopes, which were stored on silica crystals to keep samples dry. Vouchers
125 were deposited at the Intermountain Herbarium (UTC); *P. cusickiana* var. *nevadensis*
126 voucher specimens collected from Mt. Washington were additionally deposited at
127 the Great Basin National Park herbarium.

128 Because past research has shown variable relations between *P. capillaris*
129 (Holmgren and Holmgren 1974) and the *P. cusickiana* species complex (Kelso et al.
130 2009), we also tried to collect *P. capillaris* in the field. However, we were unable to
131 locate any *P. capillaris* individuals in the Ruby Mountains: at one location suggested by
132 past herbaria data, a population of *P. parryi* was found instead. To compensate, two *P.*
133 *capillaris* samples were sourced from herbaria (see Appendix I).

134 Leaf tissue from 89 samples—87 silica-dried field collections representing all
135 samples sites, and two herbarium specimens of *P. capillaris*—were placed into
136 QIGAEN Collection Microtubes (catalog number 19560) and sent to University of
137 Wisconsin-Madison Biotechnology Center, for DNA extraction, library prep, and
138 DNA sequencing (described below). Seven replicate samples were also included to

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

139 assess the quality of sequencing results, and were distributed across all four *P.*
140 *cusickiana* varieties, as well as *P. parryi*.
141 ***DNA Extraction***—DNA was extracted using the QIAGEN Dneasy mericon 96
142 QIAcube HT Kit. DNA was quantified using the Quant-iT™ PicoGreenR© dsDNA kit
143 (Life Technologies, Grand Island, New York).
144 ***Library Prep and Sequencing***—Libraries were prepared following Elshire et al.
145 2011. *ApekI* (New England Biolabs, Ipswich, Massachusetts) was used to digest 100 ng
146 of DNA. Following digestion, Illumina adapter barcodes were ligated onto DNA
147 fragments using T4 ligase (New England Biolabs, Ipswich, Massachusetts). Size
148 selection was run on a PippinHT (Sage Science, Inc., Beverly, Massachusetts) to subset
149 samples down to 300—450 bp fragments, after which samples were purified using a
150 SPRI bead cleanup. To generate quantities required for sequencing, adapter-ligated
151 samples were pooled and then amplified, and a post-amplification SPRI bead cleanup
152 step was run to remove adapter dimers. Final library qualities were assessed using the
153 Agilent 2100 Bioanalyzer and High Sensitivity Chip (Agilent Technologies, Inc., Santa
154 Clara, California), and concentrations were determined using the Qubit© dsDNA HS
155 Assay Kit (Life Technologies, Grand Island, New York). Libraries were sequenced on an
156 Illumina NovaSeq 6000 2x150.
157 ***Data Processing***—Raw FASTQ data files were demultiplexed and processed
158 using steps 1—7 of the *ipyrad* software, version 0.9.31 (Eaton and Overcast 2020).
159 Single nucleotide polymorphisms (SNPs) recognized by *ipyrad* were used as the basis for
160 variation between individuals for downstream analyses, and libraries were assembled *de*
161 *novo*. All *ipyrad* and STRUCTURE parameter files, as well as R scripts used for analysis

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

162 and data visualization, can be found on GitHub (github.com/akoontz11/Primula/) and in
163 the Supplementary Materials (SupplementalMaterials1.zip). Raw, demultiplexed
164 sequencing data can be accessed on the NCBI Sequence Read Archive (SRA; accession
165 number PRJNA705310).

166 COMPLEX-WIDE GENOMIC SURVEY—For our complex-wide genomic survey, we
167 ran *ipyrad* twice: we used the results from our initial run to confirm sequencing
168 consistency for replicate samples, and to identify samples with low coverage. For
169 both runs, demultiplexed sequences were paired and merged, and low quality bases,
170 adapters, and primers were filtered prior to SNP calling. Default values were used
171 for the *ipyrad* parameters in these steps, as well as for the clustering threshold
172 (clust_threshold; 0.85) and minimum sequencing depth (mindepth_statistical; 6)
173 parameters.

174 For our initial run, we specified a minimum number of samples per locus
175 (min_samples_locus) parameter of 10, in order to obtain loci shared between two to three
176 sample locations for any variety. Using the results from this run, we used the Python
177 script vcf2Jaccard.py to compare samples with replicates by calculating the mean Jaccard
178 similarity coefficients between all samples. We found that all replicates matched highly
179 with their corresponding samples (Fig. S1).

180 After merging replicates and removing low coverage (generally, less than 30
181 loci in the final assembly) samples from the dataset, 82 of our 87 original samples
182 remained for our complex-wide analysis. We reran *ipyrad* (steps 1-7) using these 82
183 samples to select for loci specific to this subset. We used a min_samples_locus
184 parameter of 32 for this second run, to match the ratio of minimum samples per

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

185 locus used in our initial run; *ipyrad* default values were used otherwise. Because
186 very low numbers of loci were retrieved for both herbarium specimens of *P.*
187 *capillaris* (possibly due to the age of these specimens), we were unable to include
188 *capillaris* in downstream clustering analyses.
189 VARIETY SPECIFIC CLUSTERING—In addition to our complex-wide survey, we
190 were interested in exploring population structure within variety *maguirei* which
191 could not be resolved using genetic loci shared across all species complex members.
192 To do so, we ran *ipyrad* on just the 18 *maguirei* samples used in our complex-wide
193 survey. Because five samples from each of the upper Logan canyon sampling sites
194 were included in our *ipyrad* assembly, we specified a min_samples_locus parameter
195 of 5; *ipyrad* default parameter values were used otherwise.

196 ***Population Analyses***

197 STRUCTURE—To visualize relations between complex members across their
198 geographic range, and to determine the number of identifiable genetic clusters
199 within the complex, we used the program STRUCTURE version 2.3 (Pritchard et al.
200 2000). STRUCTURE uses Bayesian clustering analysis to probabilistically assign
201 individuals to one or more of K source populations, where the loci within each
202 population are assumed to be in Hardy-Weinberg proportions and linkage
203 equilibrium. For all STRUCTURE runs, we used a burnin length of 50,000, and
204 100,000 MCMC reps after burnin. For our complex-wide survey, we ran STRUCTURE
205 for K values of 2—16, with 50 replicates per K value. For our *maguirei*-only
206 analyses, we ran STRUCTURE for K values of 2—6, with 50 replicates per K value.
207 We used the CLUMPAK server (Kopelman et al. 2015) to summarize results across

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

208 replicates for each K value, and to build STRUCTURE plots.

209 For all of our STRUCTURE analyses, we ran the Evanno et al. (2005) method
210 (which identifies the greatest ΔK value) and the method described in the STRUCTURE
211 manual (Pritchard et al. 2000, which identifies the K value with the greatest likelihood) to
212 determine an "optimal" K value. Given the difficulties in inferring an unambiguous
213 number of genetic clusters from any given set of populations (Novembre 2016; Pritchard
214 et al. 2000), we also examined STRUCTURE outputs within a range of K values, to
215 determine which value of source populations best illustrated divisions within the species
216 complex.

217 DISCRIMINANT ANALYSIS OF PRINCIPAL COMPONENTS—In addition to STRUCTURE,
218 we analyzed the results of our complex-wide survey using Discriminant Analysis of
219 Principal Components (DAPC; Jombart et al. 2010) in the package *adegenet* in R version
220 3.6.3 (R Core Team, 2020). DAPC is a statistical technique designed to accommodate the
221 size of genomic data sets and capable of differentiating within-group variation from
222 between-group variation. SNP data is first transformed using a principal components
223 analysis (PCA), and then k-means clustering is run to generate models and likelihoods
224 corresponding to each number of population clusters. The best-fitting model, and so the
225 best-supported number of populations, is assessed using the models' Bayesian
226 Information Criterion (BIC) scores. We chose to utilize DAPC in addition to
227 STRUCTURE to visualize population clusters in a PCA format, and to determine
228 whether the supported number of populations was congruent between methods, indicating
229 a more robust determination of the number of species contained within the complex
230 (Carstens et al. 2013).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

231 F_{ST} ESTIMATES—Because we wanted to measure the extent of genetic variance
232 within the groups analyzed, we used the VCFtools software (Danecek et al. 2011) to
233 generate weighted F_{ST} estimates (Weir and Cockerham 1984). We generated an F_{ST}
234 estimate for our complex-wide analysis (across all populations of all *P. cusickiana*
235 varieties) as well as for the samples included in our variety *maguirei*-only analysis.

236 RESULTS

237 ***Complex-Wide Genomic Survey***—We retrieved, on average, 2.04 x 10^6 reads
238 per sample, and our complex-wide *ipyrad* run identified 1,277 loci that were used in
239 our subsequent STRUCTURE analysis. Using the Evanno et al. (2005) method
240 yielded an optimal K value of K = 5; using the method described in the STRUCTURE
241 manual (Pritchard et al. 2000) identified the K value with the greatest likelihood as
242 K = 14. Based on our visualization of the STRUCTURE results for values ranging
243 from K = 2—16 (Figs. S2 - S4), we determined K = 7 to be the most biologically
244 relevant K value (Fig. 2). At this level of source populations, varieties *domensis* and
245 *maguirei* are clearly delineated, variety *nevadensis* shows distinctions between its
246 two populations, and variety *cusickiana* is split into three groups composed of
247 populations from the Snake River Plain in Idaho (SRP), Nevada (Jarbidge), and
248 Oregon (Owyhee). Since higher K values emphasize the divisions seen at this level,
249 and further subdivide isolated populations of varieties *cusickiana* and *nevadensis*, K
250 = 7 is a conservative estimate which reflects the strong divisions within this
251 complex while allowing for further distinctions between unique populations to be
252 made in light of more evidence in the future.

253 Our DAPC analysis revealed that the greatest supported number of clusters

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

254 (i.e. the value with the lowest BIC score) was eleven (data not shown)—a value
255 incongruent with our STRUCTURE results, suggesting that boundaries within this
256 complex are elaborate. However, at this level of genetic clusters, several groups
257 were quite small (consisting of only one or two samples), and groupings were
258 incoherent within the spatial distribution of populations. To provide a clearer
259 comparison to our STRUCTURE results, and to examine relations strictly within the
260 species complex, we removed *P. parryi* outgroup samples from our dataset (because
261 these were separate from all species complex samples in preliminary analyses) and
262 ran our DAPC with a specification of six clusters (Fig. 3). At this level of clustering,
263 the population of *nevadensis* in the Snake Range of Great Basin National Park
264 (GRBA) is shown as a unique cluster, while the *nevadensis* population further south
265 in the Grant Range groups with the *cusickiana* population sampled from Oregon
266 (Owyhee). Variety *domensis* is a unique cluster which groups closely to both of
267 these. Thus, while neither our STRUCTURE analysis nor our DAPC point to an
268 unambiguous number of "true" genetic clusters, both suggest that the current
269 varietal classification is inexact. The extreme level of divergence between the sky
270 island populations in this species complex is reflected not only in our clustering
271 analyses, but also in our relatively large F_{ST} estimate across all complex populations,
272 which was 0.72. Figure 4 illustrates proportions of sample membership to clusters
273 based on our STRUCTURE analysis at K=7 for all populations in their geographic
274 context across the Great Basin.
275 ***Variety Specific Clustering***—In our complex-wide analysis, all *maguirei* samples
276 grouped as a single cluster, distinct from all other populations of all other varieties,

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

277 indicating that neither Logan Canyon population is more closely related to any
278 populations of another variety. Even at values of K = 16, the upper and lower Logan
279 Canyon populations of *maguirei* were not resolved from one another.
280 However, reducing our sample set to only *maguirei* samples allowed us to retain loci
281 informative to this variety but unshared with other complex member populations.
282 Our *maguirei*-only *ipyrad* run generated an assembly with 68,492 loci, indicating a
283 large number of loci specific to *maguirei* and not shared with the wider species
284 complex. To speed up processing times, we ran STRUCTURE on a 17,988 loci subset
285 of *maguirei*-specific markers. Using the CLUMPAK server, we found optimal K values
286 of K = 4 (using the Evanno method) and K = 3 (using the likelihood method
287 described in the STRUCTURE manual). Figure 5 shows the STRUCTURE plot at K = 3,
288 which resolves similar groupings of *maguirei* populations supported in Bjerregaard
289 and Wolf (2004), and the distinctions between upper and lower canyon populations.
290 We also estimated an F_{ST} value of 0.33 among these three populations, which is
291 comparable to previous estimates in Bjerregaard and Wolf (2004).

292 ## DISCUSSION

293 Analysis of RADseq data from *Primula cusickiana* complex members
294 demonstrates that the disjunct geographical distribution of populations across the Great
295 Basin is reflected by pronounced genomic divergences. While the results of our
296 clustering analyses coincide with the current varietal classifications, there are notable
297 exceptions. Distinctions between isolated populations within varieties, as well as
298 similarities between neighboring populations of different varieties, can be observed in our
299 STRUCTURE plots for low K values (i.e. ranging from 2—6; see FIGS. S2-S4). For

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

300 instance, we found Mt. Washington *nevadensis* populations to be admixed, with segments
301 coming from *domensis* to the east and (to a lesser extent) Grant Range *nevadensis*
302 populations to the south. This is in accordance with analysis of AFLP and chloroplast
303 DNA from the *Primula* section Parryi, which found these two varieties to be extremely
304 close (Kelso et al. 2009).

305 Our results also suggest a more nuanced understanding of variety *cusickiana*.
306 Populations of this variety are split into distinct genomic clusters in our analysis, with
307 Jarbidge (Nevada) and Owyhee (Oregon) populations appearing unique from each other
308 and the remaining Snake River Plain (SRP) populations in Idaho. That these distinctions
309 are seen in both our STRUCTURE and DAPC analyses imply the robustness of this
310 result. Given the relatively wide distribution of this variety (growing in moist soils at
311 lower elevations than other complex members), our findings of genomic divergence
312 between its populations is noteworthy, and support past evidence of phenotypic
313 differences in different portions of its range. For instance, past morphological research of
314 Idaho *cusickiana* populations has suggested dividing this taxa into three unique species
315 (Mansfield 1993), with Owyhee populations being classified as *P. wilcoxiana*.

316 The separation between populations within variety *cusickiana*, as well as our
317 support of past findings of significant genetic distances between the proximate
318 populations of variety *maguirei*, underscore our discovery of profound genomic
319 divergences between all members of this species complex, despite their distribution over
320 a relatively small geographic area. This trend is reflected not only in our clustering
321 analyses, but also in our weighted F_{ST} estimate of 0.72 across complex populations—a
322 high value compared to similar estimates for other plant taxa (for instance, the mean F_{ST}

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

323 for plant taxa in a meta-analysis by Leinonen et al. 2008 was calculated to be 0.24). Our
324 results therefore support the historical designation of species for these complex members,
325 rather than variety. Below, we consider how two phenomena—biogeographical trends in
326 the Great Basin, and reproductive traits specific to *Primula*—may contribute to the
327 significant divergence of these populations into distinct genomic groups.

328 ***Great Basin Sky Island Biogeography***—Members of the *P. cusickiana*
329 complex are found at relatively high elevations throughout the Great Basin. Many of
330 these are sky island locations associated with strong ecological shifts as habitat
331 transitions from lower sagebrush steppe to cooler, more forested regions
332 dominated by pinyon and juniper. Now separated by arid basins due to climatic
333 warming in the Holocene, these sky islands are understood to be the fragmented
334 remnants of a continuous region of cool, moist habitat which once extended across
335 the Great Basin (Thompson and Mead 1982). This has led to their characterization
336 as refugia for various taxa—particularly mammals (Brown 1971; Badgley et al.
337 2014), but also butterflies (Boggs and Murphy 1997) and plants (Harper et al. 1978;
338 Nowak et al. 1994; Charlet 2007). Additionally, in conjunction with climatic niche
339 preferences, complex varieties *maguirei*, *domensis*, and *nevadensis* are found on the
340 cliffs and crevices of exclusively limestone substrates. While it's unclear whether
341 these habitats are tied to mineral or pH constraints, or simply reflect preferences for
342 moisture-retentive substrates, edaphic heterogeneity is known to contribute to
343 plant speciation and biodiversity, both globally (Hulshof and Spasojevic 2020) and
344 within the Great Basin (e.g. de Queiroz et al. 2012). Therefore, allopatry across
345 relatively similar climatic and edaphic niches seems to contribute to the genomic

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

346 divergences in *P. cusickiana*'s populations—a trend observed in other sections of
347 Primulaceae, as well (Boucher et al. 2016).
348 However, it has also been noted that many species distribution patterns
349 among Great Basin mountaintops do not follow a strictly island biogeographical
350 model (Lawlor 1998), in that neither island surface area nor proximity to
351 "mainland" source populations (typically identified as the western Sierra Nevadas
352 or eastern Rocky Mountains) is predictive of species abundance (Fleishman et al.
353 2001). And in some taxa, there is evidence for regular, modern dispersal between
354 Great Basin ranges (Floyd et al. 2005). An alternative scenario is that this complex
355 has followed what has been described as an "expanding-contracting archipelago"
356 (ECA) model, in response to Quaternary glacial cycles (DeChaine and Martin 2005a).
357 The ECA model has been used to describe the divergence between Rocky Mountain
358 sky island plant taxa (Dechaine and Martin 2005b; Hodel et al. 2021), and provides a
359 framework for explaining the genetic structure observed between isolated montane
360 populations on a broad spatial scale. In this model, populations are assumed to
361 become fragmented as they contract up-slope during warmer interglacials; during
362 glacial periods, populations expand down-slope as moist, cool habitat becomes
363 widespread, leading to hybrid zones and possible admixture. Given the degree of
364 fragmentation between *P. cusickiana*'s populations in today's climate (which
365 resembles past interglacial periods), and the admixture between the relatively
366 proximate populations of varieties *domensis* and *nevadensis* revealed in our analysis,
367 this model offers a viable explanation for the trends observed in this species
368 complex. In addition to these biogeographic patterns, the evolution of *P. cusickiana*'s

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

369 disjunct populations is simultaneously influenced on a finer spatial scale by aspects
370 particular to this species' biology.

371 ***Speciation and Heterostyly in Primula***—Recent research has shown several
372 different alpine *Primula* species complexes to contain previously undescribed
373 cryptic species, in China (Huang et al. 2019; Ren et al. 2020) and in Europe (Schorr
374 et al. 2013; Theodoridis et al. 2019). Our findings on the *P. cusickiana* species
375 complex resonate with these trends, and raise the question of what unique traits
376 *Primula* possesses which might cause such frequent diversification via allopatric
377 speciation. The authors of a study examining the *P. merrilliana* species complex in
378 China (He et al. 2021) argue that heterostyly—a widespread breeding system in
379 angiosperms to promote outcrossing—may be a driving force leading to speciation.
380 In heterostyly, "pin" and "thrum" floral morphologies prevent self-fertilization via
381 insect pollination (Darwin 1897), and are associated with a sporophytic-
382 incompatibility system which follows a Mendelian pattern of inheritance (Li et al.
383 2016). In *P. merrilliana*, the efficacy and prevalence of heterostyly and self-
384 incompatibility varies across populations, which has possibly led to the divergence
385 between distylous and homostylous populations and, ultimately, speciation.

386 While the presence of heterostyly has been observed in *nevadensis*
387 (Holmgren 1967) and in populations of *cusickiana* and *domensis* (pers. obs.), the
388 extent of distyly in a population has only been well documented in *maguirei*, who's
389 upper and lower canyon populations have a pin:thrum morphology ratio of about
390 1:1 (Davidson et al. 2014). This implies that in scenarios of legitimate xenogamy, in
391 which morphs of one type only mate with morphs of the opposite type, only half of

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

392 the total population is available as a potential mate for any distylous individual.
393 While this reduction in effective population size would seem to increase the
394 strength of genetic drift, and possibly the fixation of deleterious alleles, these
395 negative effects are potentially counterbalanced by the genetic advantages of
396 outcrossing. This net benefit of heterostyly is supported by findings in de Vos et al.
397 (2014), in which phylogenetic techniques were used to demonstrate that the
398 presence of heterostyly in *Primula* leads to greater diversification via decreased
399 extinction, in the long-term, compared to non-heterostylous clades of Primulaceae.
400 Simultaneously, the loss of heterostyly and subsequent self-compatibility may lead
401 to rapid speciation in the short-term. Observation of distylous morph ratios in other
402 *P. cusickiana* varieties and populations, and changes in these ratios between
403 proximate populations, would help to determine if these dynamics are driving the
404 divergences we see at the species complex level.
405 ***Conclusion***—The results of our genomic survey of *Primula cusickiana* fit into
406 a wider trend demonstrating abundant allopatric speciation despite little niche
407 divergence in other alpine *Primula* species complexes. Our findings support the
408 historical classification of each of these complex members as unique species, rather
409 than the varietal classification taken in Holmgren and Kelso (2001). Furthermore,
410 these results warrant a more detailed understanding of the isolated and genetically
411 unique populations in this complex (such as *cusickiana* populations in Nevada and
412 Oregon), and of the admixture observed in the populations of variety *nevadensis*.
413 Similarly, updated morphological comparisons between varieties, as well as
414 observations into the levels of heterostyly in disjunct populations, would offer a

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

415 clearer understanding of the mechanisms of speciation occurring within this
416 complex. Finally, the endemic species with narrow niches included in this study,
417 such as *P. cusickiana* var. *maguirei*, but also *nevadensis*, *domensis*, and the sister
418 species *P. capillaris*, warrant concern of extinction, and more work needs to be done
419 to better understand the breeding limitations faced by each of these taxa and what
420 can be done to ensure their survival in an increasingly arid Great Basin.

421 ## ACKNOWLEDGMENTS

422 his research was supported by funds provided by the Margaret Williams
423 Research Grant offered by the Native Nevada Plant Society, the Lawrence Piette
424 Graduate Scholarship offered by the USU College of Science, the Dr. Ivan J. Palmblad
425 Graduate Research Award offered by the USU Department of Biology, and the
426 Graduate Research Award offered by the USU Ecology Center. WDP and the Pearse
427 Lab are funded by NSF EF-1802605, NSF ABI-1759965, and UKRI-NERC
428 NE/V009710/1.

429 The authors would like to thank Dr. Carol Rowe for her help with the genetic
430 analyses, her Python script used to calculate Jaccard similarities from SNP data, and her
431 Python script for creating the map. Thanks also to Dr. Barbara Ertter and especially Dr.
432 Don Mansfield at the College of Idaho, for their assistance in collecting variety
433 *cusickiana*. Trish Winn and Jennifer Lewihnsohn in the United States Forest Service
434 provided the collection permit for variety *maguirei* from the Uinta-Wasatch-Cache
435 National Forest; Todd Stefanic and Gretchen Baker with the National Park Service
436 coordinated collection from Craters of the Moon National Monument and Great Basin
437 National Park, respectively. Michael Piep, Elizabeth Makings, and Jerry Tiehm allowed

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

438 for *Primula* specimens to be sampled from the Intermountain Herbarium, Arizona State
439 Vascular Plant Herbarium, and University of Nevada, Reno Herbarium, respectively.
440 Thanks also to Kris Valles at the Intermountain Herbarium. Dr. Leila Shultz assisted with
441 identification of the Owyhee *cusickiana* population. The authors would also like to thank
442 Jean Howerton, Buddy Smith, and Noel and Pat Holmgren. Finally, all collections were
443 made on the ancestral lands of the Western Shoshone, Eastern Shoshone, Shoshone-
444 Bannock, Southern Paiute, Goshute, and Nez Perce Native American tribes.

445 ## AUTHOR CONTRIBUTIONS

446 AK determined sample locations, performed the majority of sample collection, and ran
447 genetic analyses. WDP contributed to study design and assisted with genetic analyses and
448 manuscript writing. PW guided study design and assisted with genetic analyses, sample
449 collection, and manuscript writing.

450

451 ## LITERATURE CITED

452 Badgley, C., Smiley, T. M., & Finarelli, J. A. (2014). Great Basin mammal diversity in
453 relation to landscape history. *Journal of Mammalogy*, *95*(6), 1090–1106.
454 Bjerregaard, L., & Wolf, P. G. (2008). Strong genetic differentiation among neighboring
455 populations of a locally endemic primrose. *Western North American Naturalist*,
456 *68*(1), 66–75.
457 Boggs, C. L., & Murphy, D. D. (1997). Community composition in mountain
458 ecosystems: climatic determinants of montane butterfly distributions. *Global*
459 *Ecology and Biogeography Letters*, *6*(1), 39–48.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

460 Boucher, F. C., Zimmermann, N. E., & Conti, E. (2016). Allopatric speciation with little
461 niche divergence is common among alpine Primulaceae. *Journal of*
462 *Biogeography*, *43*(3), 591–602.
463 Brown, J. H. (1971). Mammals on mountaintops: nonequilibrium insular biogeography.
464 *The American Naturalist*, *105*(945), 467–478.
465 Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species
466 delimitation. *Molecular Ecology*, *22*(17), 4369–4383.
467 Charlet, D. A. (2007). Distribution patterns of Great Basin conifers: Implications of
468 extinction and immigration. *Aliso: A Journal of Systematic and Floristic Botany*,
469 *24*(1), 31–61.
470 Coates, D. J., Byrne, M., & Moritz, C. (2018). Genetic diversity and conservation units:
471 Dealing with the species-population continuum in the age of genomics. *Frontiers*
472 *in Ecology and Evolution*, *6*, 165.
473 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A.,
474 Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R.,
475 & 1000 Genomes Project Analysis Group. (2011). The variant call format and
476 VCFtools. *Bioinformatics* , *27*(15), 2156–2158.
477 Darwin, C. (1877). *The Different Forms of Flowers on Plants of the Same Species*. John
478 Murray, London.
479 Davidson, J. B., Durham, S. L., & Wolf, P. G. (2014). Breeding system of the threatened
480 endemic *Primula cusickiana* var. *maguirei* (Primulaceae). *Plant Species Biology*,
481 *29*(3).
482

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

483 de Queiroz, T. F., Baughman, C., Baughman, O., Gara, M., & Williams, N. (2012).
484 Species distribution modeling for conservation of rare, edaphic endemic plants in
485 White River Valley, Nevada. *Natural Areas Journal*, *32*(2), 149–158.
486 de Vos, J. M., Hughes, C. E., Schneeweiss, G. M., Moore, B. R., & Conti, E. (2014).
487 Heterostyly accelerates diversification via reduced extinction in primroses.
488 *Proceedings of the Royal Society B*, *281*(1784).
489 DeChaine, E. G., & Martin, A. P. (2005a). Historical biogeography of two alpine
490 butterflies in the Rocky Mountains: broad-scale concordance and local-scale
491 discordance. *Journal of Biogeography*, *32*(11), 1943–1956.
492 Dechaine, E. G., & Martin, A. P. (2005b). Marked genetic divergence among sky island
493 populations of *Sedum lanceolatum* (Crassulaceae) in the Rocky Mountains.
494 *American Journal of Botany*, *92*(3), 477–486.
495 Eaton, D. A. R., & Overcast, I. (2020). ipyrad: interactive assembly and analysis of
496 RADseq datasets. *Bioinformatics, 36*(8)*,* 2592–2594.
497 Elshire, R. J., Glaubitz, J. C., Sun, Q., Pol, J. A., Kawamoto, K., Buckler, E. S., &
498 Sharon, E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach
499 for high diversity species. *PloS One*, *6*(e19379).
500 Fish and Wildlife Service, Department of the Interior. (1985). *Endangered and*
501 *Threatened Wildlife and Plants; Final Rule to Determine Primula Maguirei*
502 *(Maguire Primrose) To Be a Threatened Species* (Vol. 50, pp. 49–52).
503 Fleishman, E., Austin, G. T., & Murphy, D. D. (2001). Biogeography of Great Basin
504 butterflies: revisiting patterns, paradigms, and climate change scenarios.
505 *Biological Journal of the Linnean Society*, *74*(4), 501–515.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

506 Floyd, C. H., Van Vuren, D. H., & May, B. (2005). Marmots on great basin
507 mountaintops: using genetics to test a biogeographic paradigm. *Ecology*, *86*(8),
508 2145–2153.
509 Gray, A. (1888). *Synoptical Flora of North America*. Ivison, Blakeman, Taylor & Co.
510 Harper, K. T., Freeman, C. D., Ostler, K. W., & Klikoff, L. G. (1978). The flora of Great
511 Basin mountain ranges: diversity sources and dispersal ecology. *Great Basin*
512 *Naturalist Memoirs*, *2*, 81–103.
513 He, X., Cao, J.-J., Zhang, W., Li, Y.-Q., Zhang, C., Li, X.-H., Xia, G.-H., & Shao, J.-W.
514 (2021). Integrative taxonomy of herbaceous plants with narrow fragmented
515 distributions: a case study on *Primula merrilliana* species complex. *Journal of*
516 *Systematics and Evolution*
517 Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in
518 divergence and speciation. *Biological Journal of the Linnean Society*, *58*(3), 247–
519 276.
520 Hodel, R. G. J., Massatti, R., Bishop, S. G. D., & Knowles, L. L. (2021). Testing which
521 axes of species differentiation underlie covariance of phylogeographic similarity
522 among montane sedge species. *Evolution*
523 Holmgren, N. H. (1967). A new species of primrose from Nevada. *Madroño*, *19*(1), 27–
524 29.
525 Holmgren, N. H., & Holmgren, A. H. (1974). Three new species from the Great Basin.
526 *Brittonia*, *26*(3), 309–315.
527 Holmgren, N. H., Holmgren, P. K., & Cronquist, A. (2005). *Intermountain flora:*
528 *vascular plants of the Intermountain West, USA. Volume two, Part B: Subclass*

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

529 *Dilleniidae*. New York Botanical Garden.
530 Holmgren, N. H., & Kelso, S. (2001). *Primula cusickiana* (Primulaceae) and its varieties.
531 *Brittonia*, *53*(1), 154–156.
532 Huang, J. (2020). Is population subdivision different from speciation? From
533 phylogeography to species delimitation. *Ecology and Evolution*, *36*, 303.
534 Huang, Y.-F., Dong, L.-N., & Xu, W.-B. (2019). *Lysimachia fanii*, a new species of
535 Primulaceae from limestone area of Guangxi, China. *PhytoKeys*, *130*, 75–84.
536 Hulshof, C. M., & Spasojevic, M. J. (2020). The edaphic control of plant diversity.
537 *Global Ecology and Biogeography*, *29*(10), 1634–1650.
538 Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal
539 components: a new method for the analysis of genetically structured populations.
540 *BMC Genetics*, *11*, 94.
541 Kass, R., & Welsh, S. (1985). New Species of *Primula* (Primulaceae) from Utah. *The*
542 *Great Basin Naturalist*, *45*(3), 548–550.
543 Kelso, S., Beardsley, P. M., & Weitemier, K. (2009). Phylogeny and Biogeography of
544 *Primula* sect. *Parryi* (Primulaceae). *International Journal of Plant Sciences*,
545 *170*(1), 93–106.
546 Lawlor, T. E. (1998). Biogeography of Great Basin mammals: paradigm lost? *Journal of*
547 *Mammalogy*, *79*(4), 1111–1130.
548 Leinonen, T., O'Hara, R. B., Cano, J. M., & Merilä, J. (2008). Comparative studies of
549 quantitative trait and neutral marker divergence: a meta-analysis. *Journal of*
550 *Evolutionary Biology*, *21*(1), 1–17.
551 Li, J., Cocker, J. M., Wright, J., Webster, M. A., McMullan, M., Dyer, S., Swarbreck, D.,

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

552 Caccamo, M., van Oosterhout, C., & Gilmartin, P. M. (2016). Genetic
553 architecture and evolution of the S locus supergene in *Primula vulgaris*. *Nature*
554 *Plants*, *2*(12), 16188.
555 Mansfield, D. (1993). The status of "*Primula wilcoxiana*" and the *Primula cusickiana*
556 complex in Southwestern Idaho. Bureau of Land Management, U.S. Department
557 of the Interior.
558 Novembre, J. (2016). Pritchard, Stephens, and Donnelly on Population Structure.
559 *Genetics*, *204*(2), 391–393.
560 Nowak, C. L., Nowak, R. S., Tausch, R. J., & Wigand, P. E. (1994). Tree and shrub
561 dynamics in northwestern Great Basin woodland and shrub steppe during the
562 Late-Pleistocene and Holocene. *American Journal of Botany*, *81*(3), 265–277.
563 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure
564 using multilocus genotype data. *Genetics*.
565 R Core Team (2020). R: A language and environment for statistical computing.
566 R Foundation for Statistical Computing, Vienna, Austria.
567 Ren, G., Mateo, R. G., Conti, E., & Salamin, N. (2020). Population genetic structure and
568 demographic history of *Primula fasciculata* in Southwest China. *Frontiers in*
569 *Plant Science*, *11*, 986.
570 Schorr, G., Pearman, P. B., Guisan, A., & Kadereit, J. W. (2013). Combining
571 palaeodistribution modelling and phylogeographical approaches for identifying
572 glacial refugia in alpine Primula. *Journal of Biogeography*, *40*(10), 1947–1960.
573 Theodoridis, S., Nogués□Bravo, D., & Conti, E. (2019). The role of cryptic diversity and
574 its environmental correlates in global conservation status assessments: Insights

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

575 from the threatened bird's□eye primrose (*Primula farinosa* L.). *Diversity and*
576 *Distributions, 25*, 1457–1471.
577 Thompson, R. S., & Mead, J. I. (1982). Late Quaternary environments and biogeography
578 in the Great Basin. *Quaternary Research*, *17*(1), 39–55.
579 Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the analysis of
580 population structure. *Evolution*, *38*(6), 1358–1370.
581 Williams, L. O. (1936). Revision of the Western Primulas. *The American Midland*
582 *Naturalist*, *17*(4), 741–748.
583 Wolf, P. G., & Sinclair, R. B. (1997). Highly differentiated populations of the narrow
584 endemic plant Maguire Primrose (*Primula maguirei*). *Conservation Biology*,
585 *11*(2), 375–381.
586 Yang, L., Kong, H., Huang, J.-P., & Kang, M. (2019). Different species or genetically
587 divergent populations? Integrative species delimitation of the *Primulina*
588 *hochiensis* complex from isolated karst habitats. *Molecular Phylogenetics and*
589 *Evolution*, *132*, 219–231.
590

591 APPENDIX 1. Voucher specimens. Order of data is as follows: Species, Voucher,
592 Herbarium. Institutional barcodes or accession numbers are included as
593 parenthetical values following the voucher, when available.
594 **Ingroup:** *Primula cusickiana* var. *cusickiana*, 25330978, Intermountain
595 Herbarium; *Primula cusickiana* var. *cusickiana*, 25330990, Intermountain
596 Herbarium; *Primula cusickiana* var. *cusickiana*, 25331045, Intermountain
597 Herbarium; *Primula cusickiana* var. *cusickiana*, 25331062, Intermountain

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

598 Herbarium; *Primula cusickiana* var. *cusickiana*, 25331021, Intermountain
599 Herbarium; *Primula cusickiana* var. *cusickiana*, 25331015, Intermountain
600 Herbarium; *Primula cusickiana* var. *cusickiana*, 25331018, Intermountain
601 Herbarium; *Primula cusickiana* var. *cusickiana*, 25331034, Intermountain
602 Herbarium; *Primula cusickiana* var. *cusickiana*, 25331004, Intermountain
603 Herbarium; *Primula cusickiana* var. *cusickiana*, 25330994, Intermountain
604 Herbarium; *Primula cusickiana* var. *cusickiana*, 25330991, Intermountain
605 Herbarium; *Primula cusickiana* var. *maguirei*, 25331026, Intermountain Herbarium;
606 *Primula cusickiana* var. *maguirei*, 25331039, Intermountain Herbarium; *Primula*
607 *cusickiana* var. *maguirei*, 25331041, Intermountain Herbarium; *Primula cusickiana*
608 var. *nevadensis*, 25331101, Intermountain Herbarium; *Primula cusickiana* var.
609 *nevadensis*, 25331106, Intermountain Herbarium; *Primula cusickiana* var.
610 *nevadensis*, 25331092, Intermountain Herbarium; *Primula cusickiana* var. *domensis*,
611 25331066, Intermountain Herbarium; *Primula cusickiana* var. *domensis*, 25331070,
612 Intermountain Herbarium; *Primula cusickiana* var. *domensis*, 25331077,
613 Intermountain Herbarium; *Primula cusickiana* var. *domensis*, 25331083,
614 Intermountain Herbarium;
615 **Outgroups:** *Primula capillaris*, 770850 (ASU0020421), Arizona State
616 University Vascular Plant Herbarium; *Primula capillaris*, 3025822 (UTC00138833),
617 Intermountain Herbarium; *Primula parryi*, 25331110, Intermountain Herbarium;
618 *Primula parryi*, 25331112, Intermountain Herbarium
619
620 FIG. 1. Four members of the *Primula cusickiana* species complex: (A) *maguirei*,

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

621 in Right Hand Fork of Logan Canyon; (B) *cusickiana*, near Cougar Point in Jarbidge,
622 Nevada; (C) *domensis*, at Notch Peak in the House Range, Utah; (D) *nevadensis*, on
623 Mount Washington in the Snake Range (Great Basin National Park), in Nevada.

624 FIG. 2. Sample STRUCTURE plots at K = 7. At this level of clustering, divisions
625 between isolated populations of variety *cusickiana* in Idaho (Snake River Plain, or
626 SRP), Nevada (Jarbidge), and Oregon (Owyhee) are clearly shown. Similarities
627 between populations of variety *nevadensis* in Great Basin National Park (GRBA) and
628 *domensis* are shown, while populations of *nevadensis* further south in the Grant
629 Range (Troy) are more distinct.

630 FIG. 3. DAPC of only *P. cusickiana* complex samples with number of genetic
631 clusters specified at 6; percentage of total variance for each PC axis shown. Similar
632 to STRUCTURE results at K = 7, this analysis shows all *maguirei* populations as
633 distinct from all other complex populations. Populations of varieties *domensis* and
634 *nevadensis* group closely with *cusickiana* population from Oregon
635 ("cusickiana_Owyhee").

636 FIG. 4. Map of sample locations with cluster membership. Sampling locations
637 are represented by pie charts indicating percentage of population membership to
638 clusters determined at K = 7 STRUCTURE clustering threshold. With exception to
639 *nevadensis*, most samples fall almost entirely within a specified cluster.

640 FIG. 5. STRUCTURE plot for *maguirei* samples at a clustering threshold of K =
641 3. While *maguirei* clustered together in the complex-wide analysis, our *maguirei*-
642 only analysis was able to reveal the Logan Canyon population divisions illustrated in
643 past studies.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

644 FIG. S1. Distribution of pairwise Jaccard similarities across all samples.

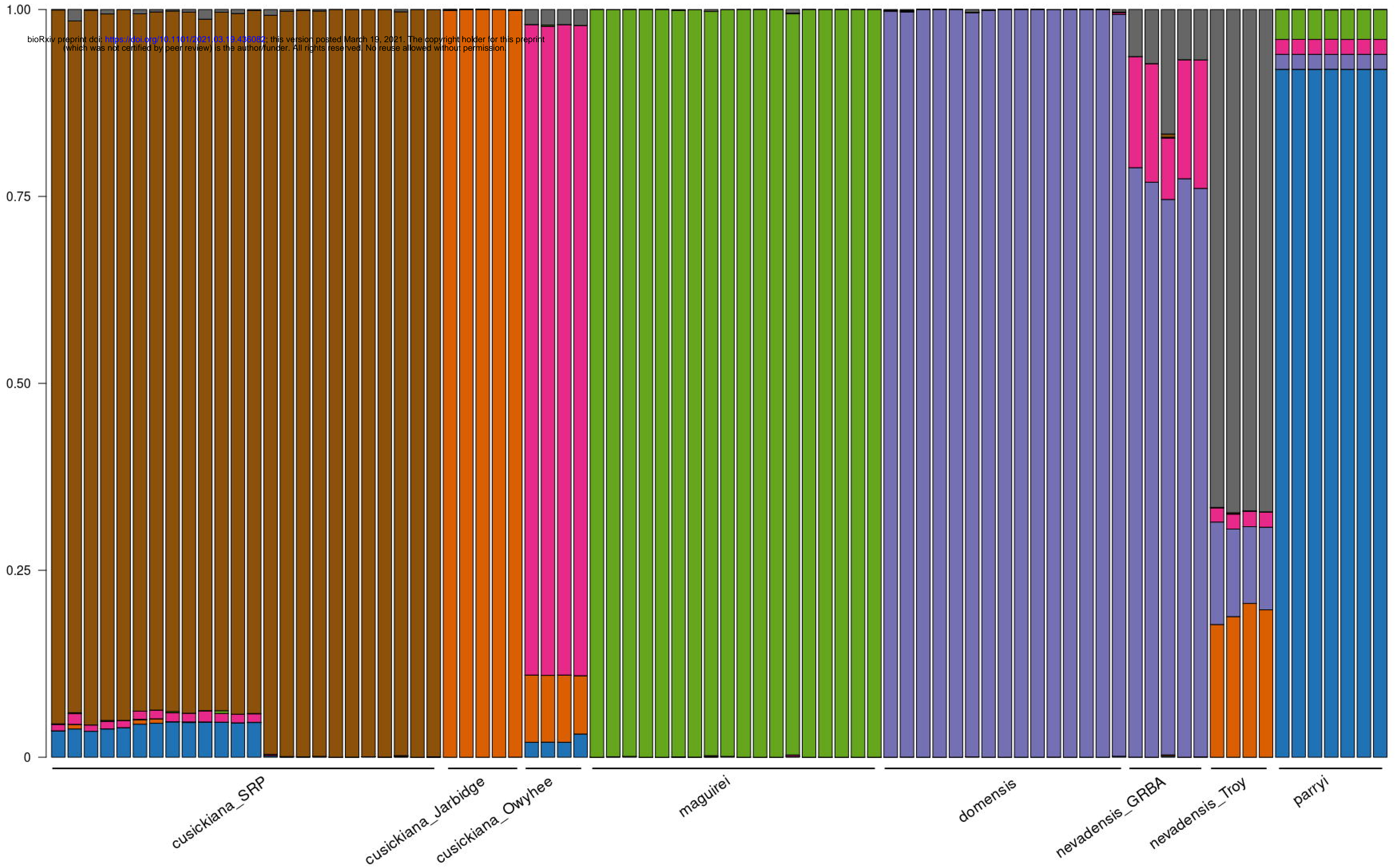
645 Similarity values of replicates are indicated by red vertical lines.

646 FIG. S2. STRUCTURE plots for all samples, K values ranging from 2 to 6.

647 FIG. S3. STRUCTURE plots for all samples, K values ranging from 7 to 11.

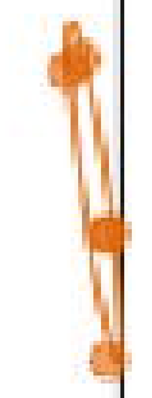
648 FIG. S4. STRUCTURE plots for all samples, K values ranging from 12 to 16.

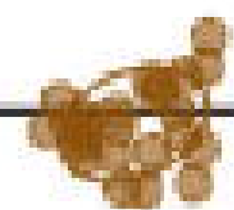
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

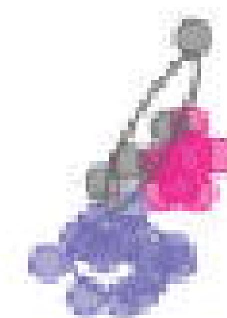


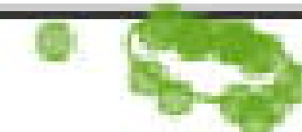
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

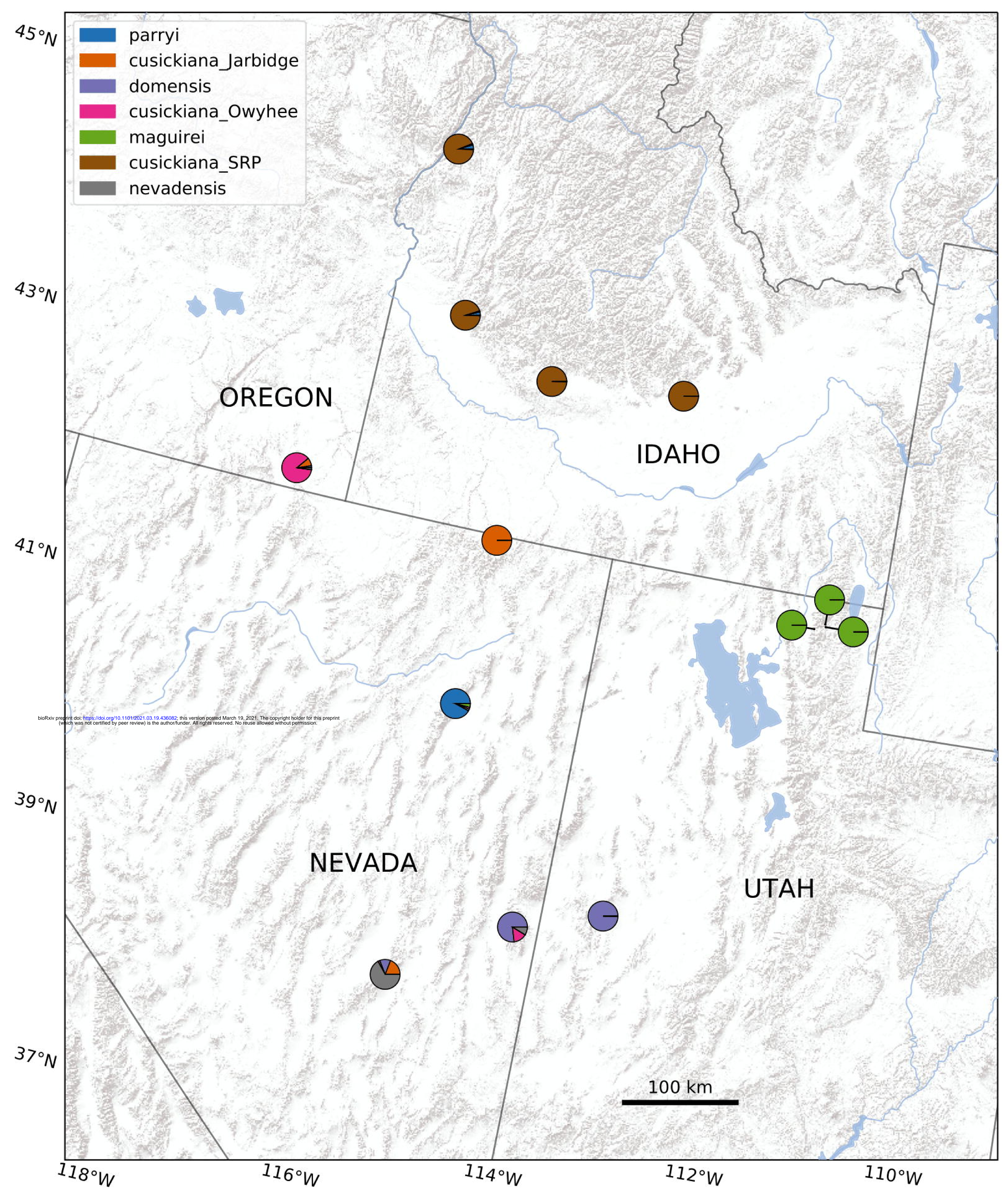
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



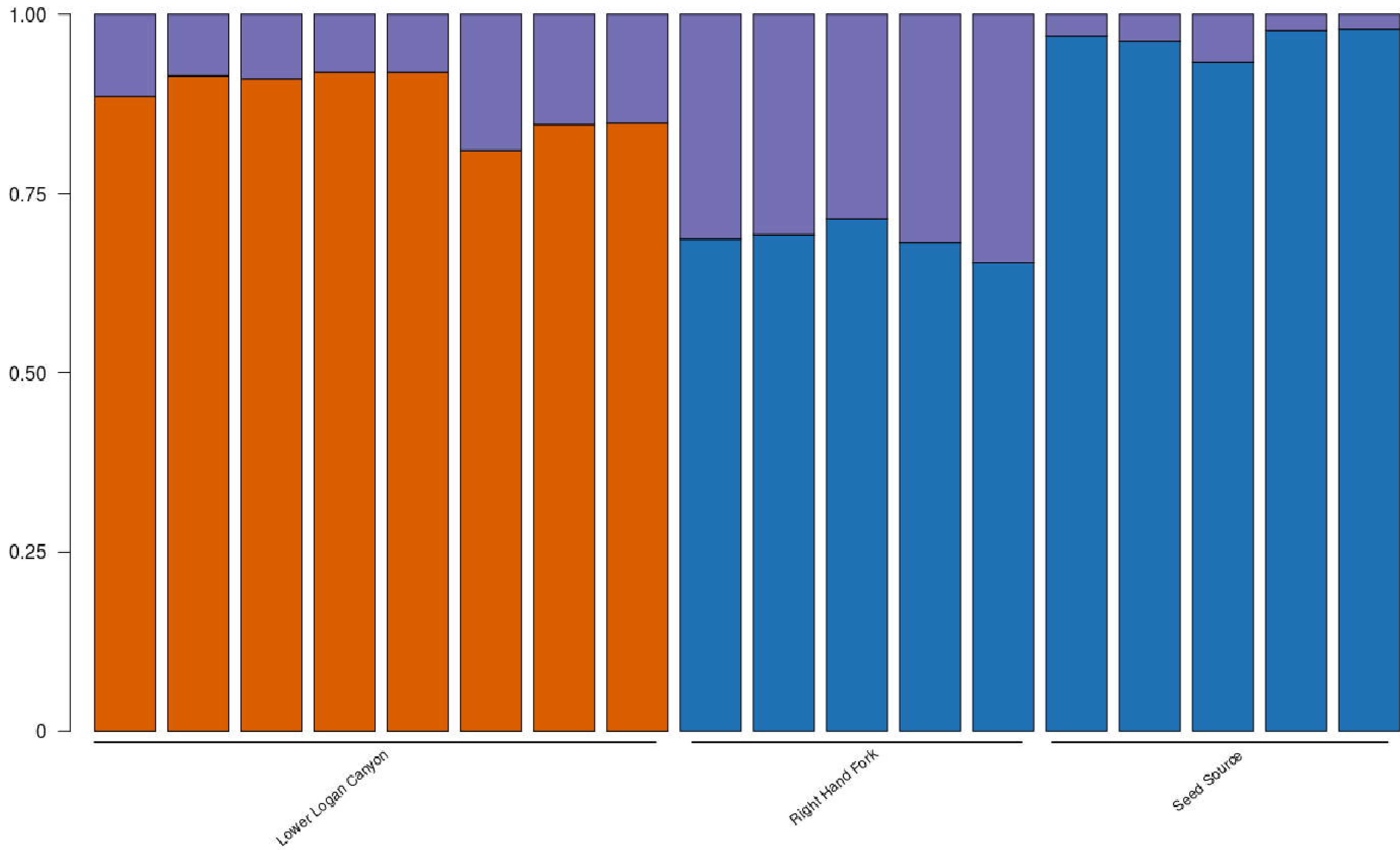




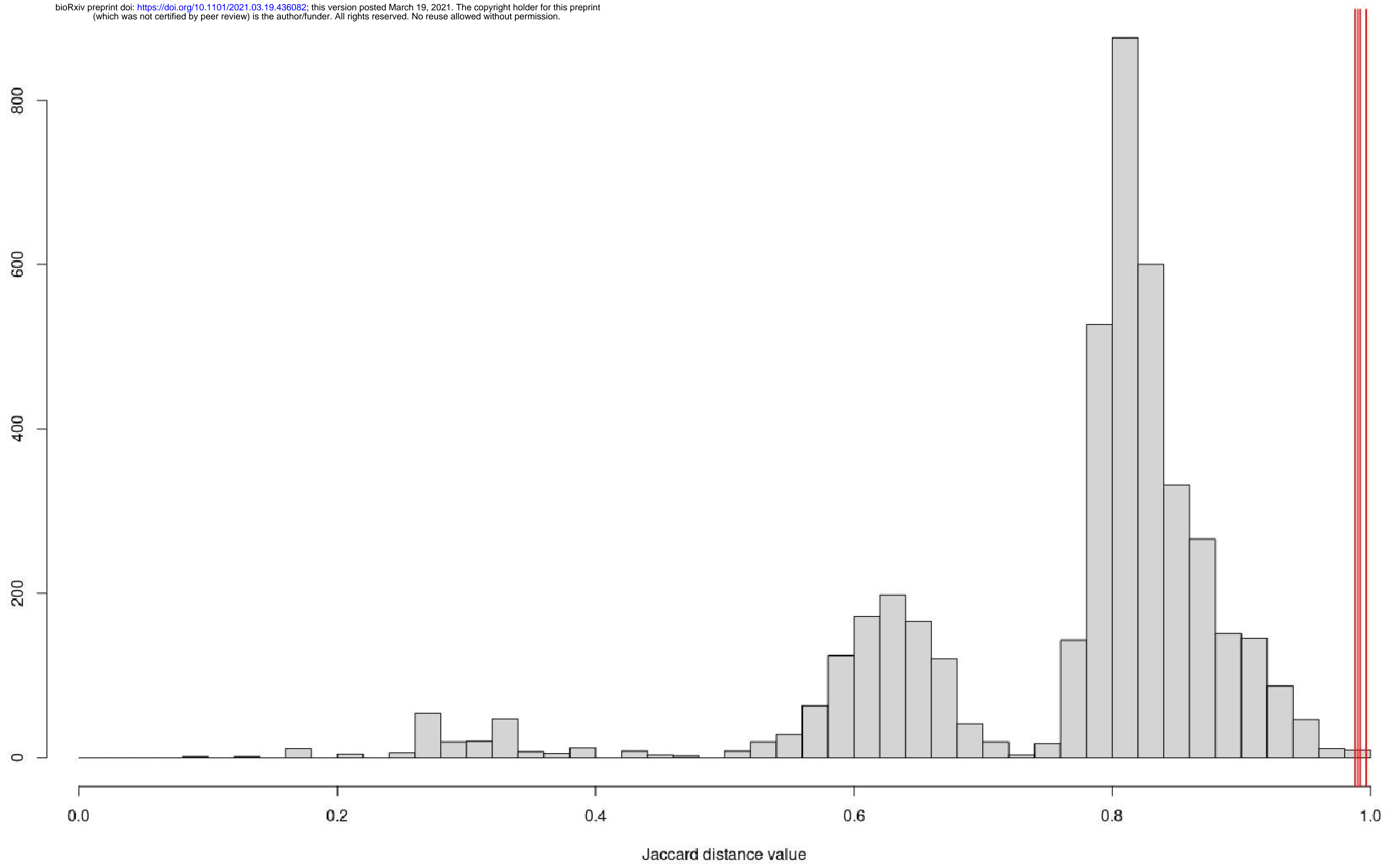




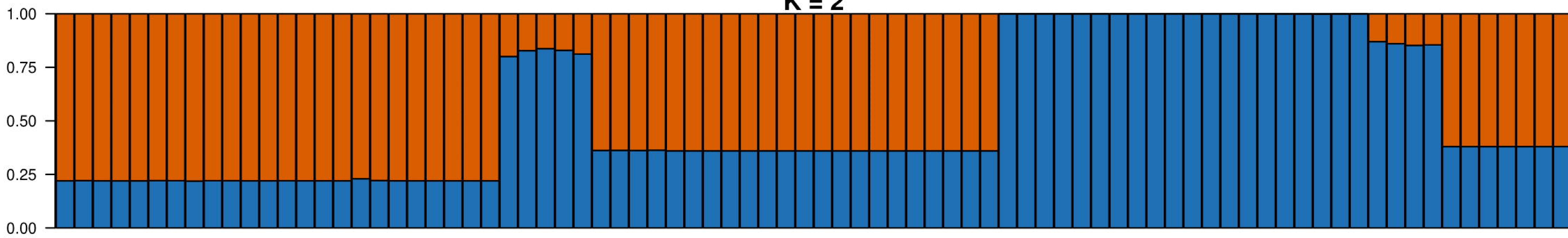
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

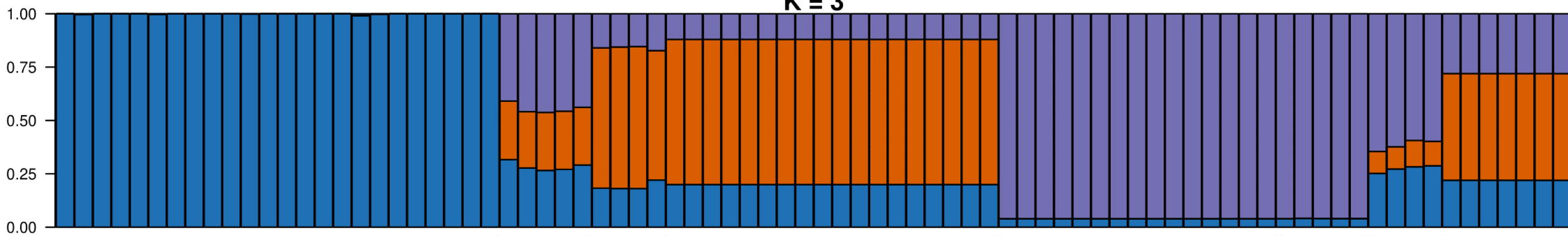

K=3
1.00
0.75
0.50
0.25
0
Lower Logan Canyon
Right Hand Fork
Seed Source

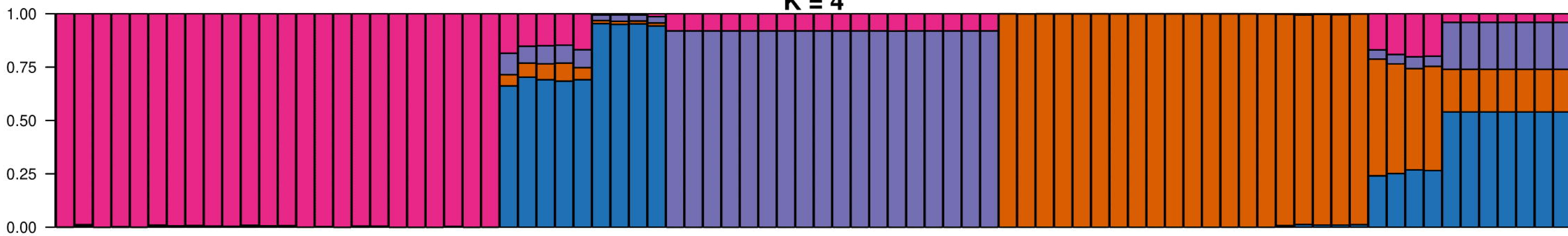
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

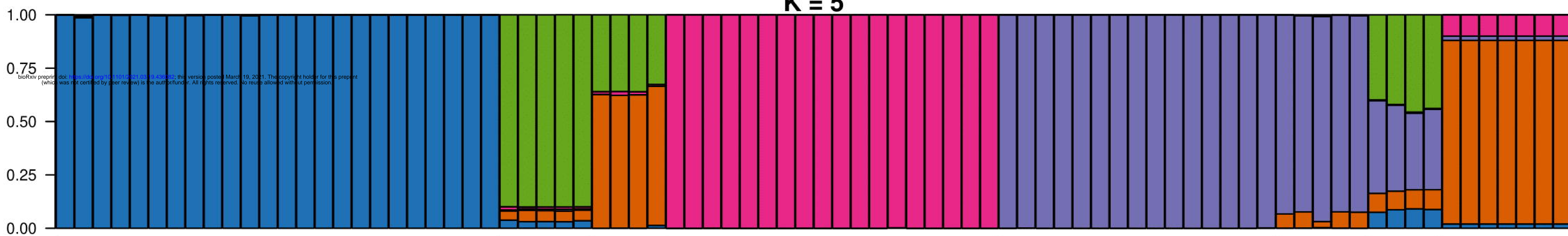


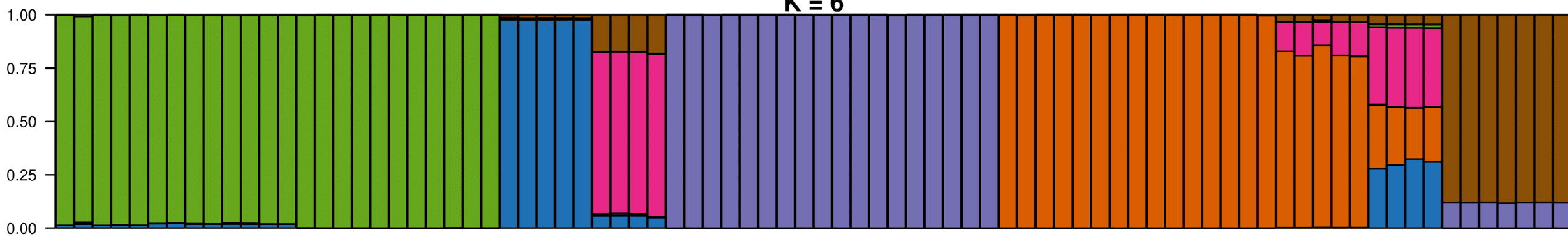
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

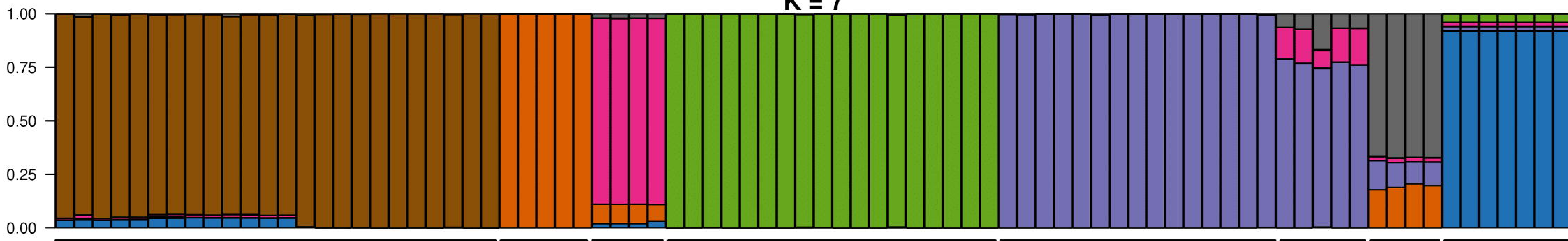


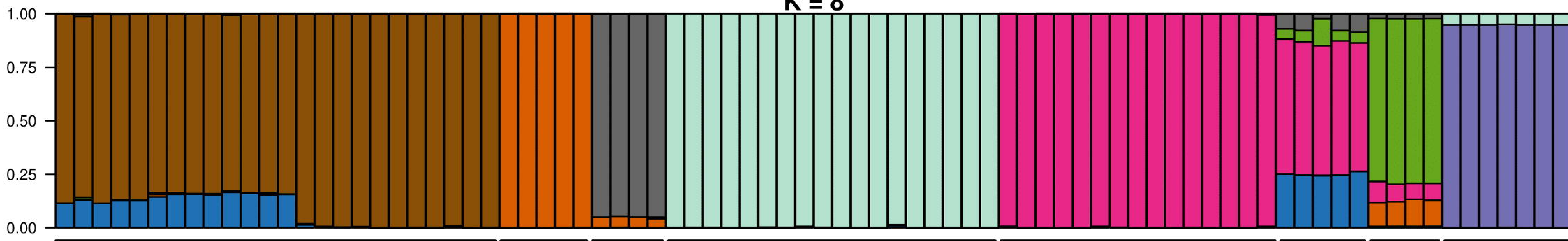


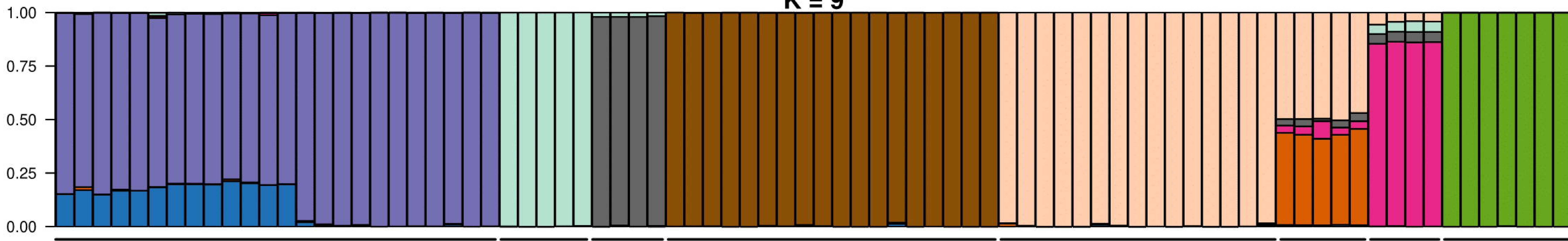


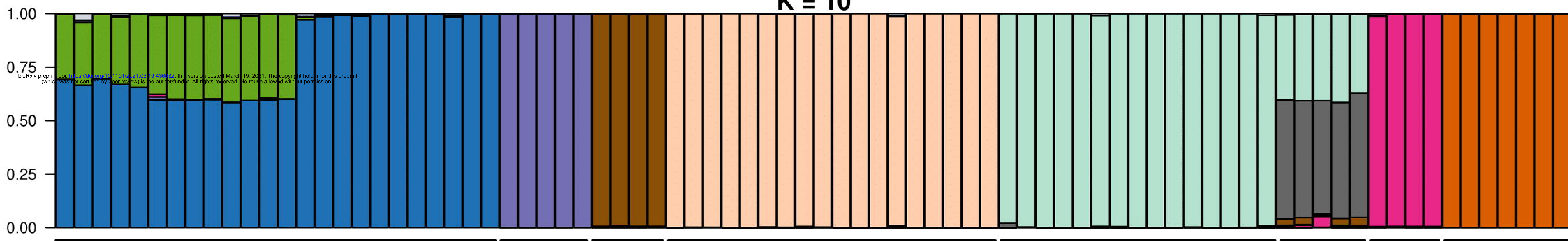




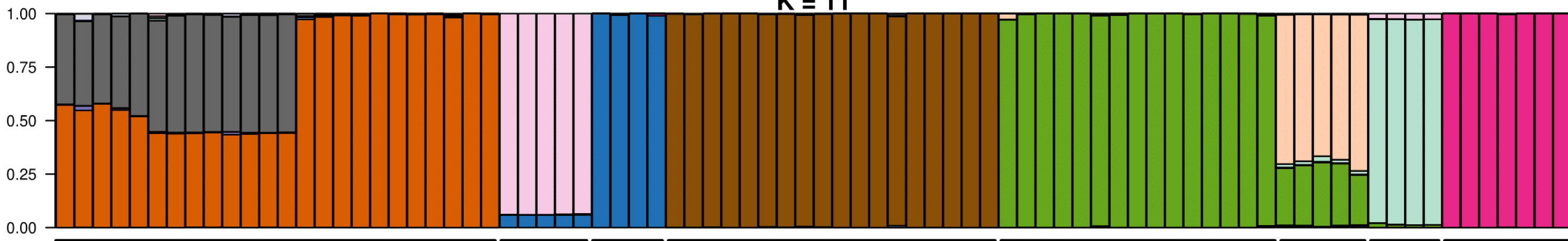


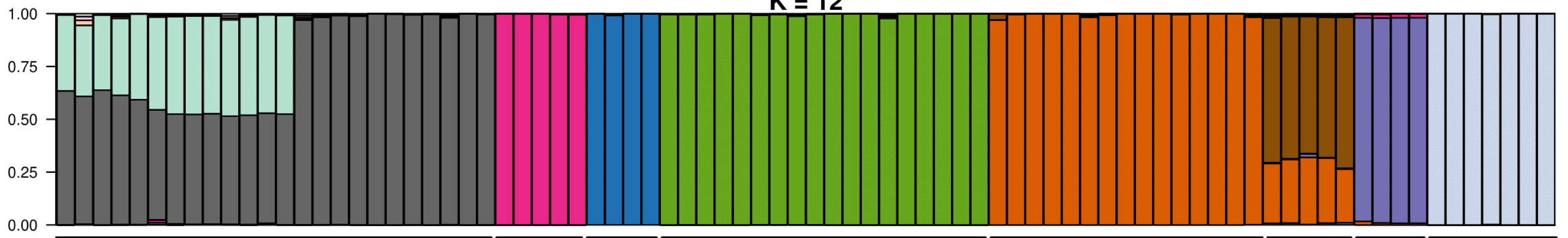


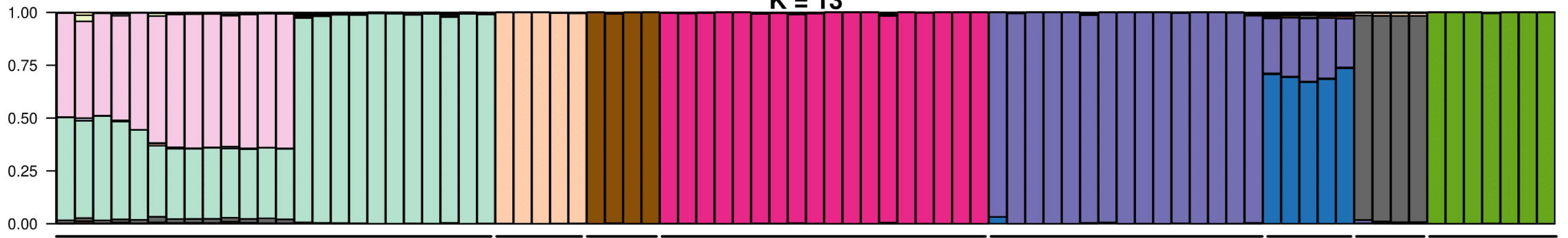


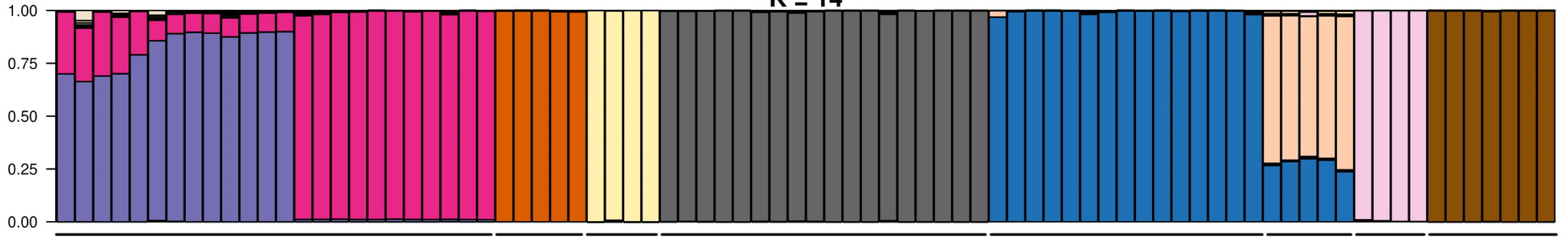


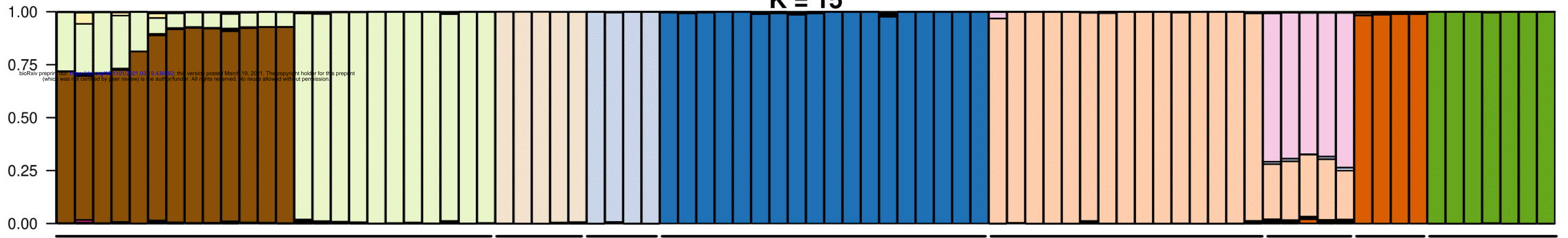
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.19.436082; this version posted March 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

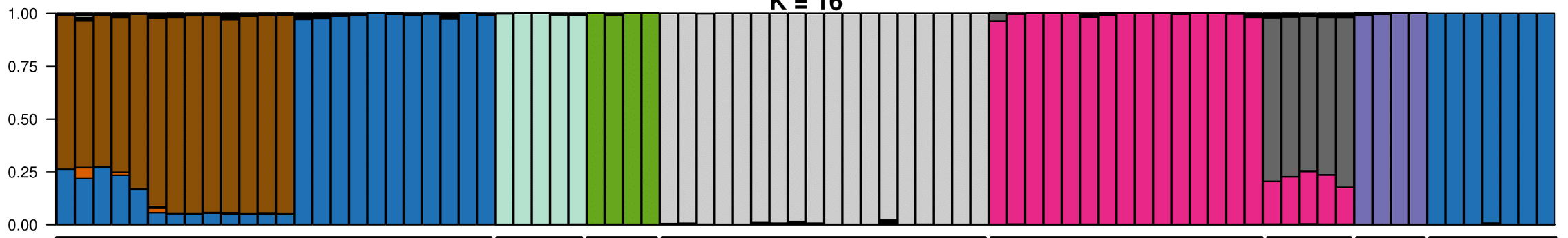



K = 12
1.00
0.75
0.50
0.25
0.00


K = 13
1.00
0.75
0.50
0.25
0.00


K = 14
1.00
0.75
0.50
0.25
0.00


K = 15
1.00
0.75
0.50
0.25
0.00


K = 16
1.00
0.75
0.50
0.25
0.00
ck_Idaho
ck_Jarbidge
ck_Owyhee
maguirei
domensis
nv_GRBA
nv_Troy
parryi