

Genetic differentiation is determined by geographic distance in

*Clarkia pulchella*

Megan Bontrager<sup>1,3</sup> and Amy L. Angert<sup>2</sup>

<sup>1</sup>Department of Botany, University of British Columbia, Vancouver, British Columbia V6T  
1Z4, Canada

<sup>2</sup>Departments of Botany and Zoology, University of British Columbia, Vancouver, British  
Columbia V6T 1Z4, Canada

<sup>3</sup>Author for correspondence (mgbontrager@gmail.com)

Running title: Isolation by distance in *C. pulchella*

## 1 Abstract

2 Both environmental differences and geographic distances may contribute to the genetic differentiation of  
3 populations on the landscape. Understanding the relative importance of these drivers is of particular inter-  
4 est in the context of geographic range limits, as both swamping gene flow and lack of genetic diversity are  
5 hypothesized causes of range limits. We investigated the landscape genetic structure of 32 populations of the  
6 annual wildflower *Clarkia pulchella* from across the species' geographic range in the interior Pacific North-  
7 west. We tested whether climatic differences between populations influenced the magnitude of their genetic  
8 differentiation. We also investigated patterns of population structure and geographic gradients in genetic  
9 diversity. Contrary to our expectations, we found an increase in genetic diversity near the species' northern  
10 range edge. We found no notable contribution of climatic differences to genetic differentiation, indicating  
11 that any processes that might operate to differentiate populations based on temperature or precipitation  
12 are not affecting the putatively neutral loci in these analyses. Rather, these results support seed and pollen  
13 movement at limited distances relative to the species' range and that this movement and the subsequent  
14 incorporation of immigrants into the local gene pool are not influenced by temperature or precipitation  
15 similarities among populations. We found that populations in the northern and southern parts of the range  
16 tended to belong to distinct genetic groups and that central and eastern populations were admixed between  
17 these two groups. This pattern could be the result of a past or current geographic barrier associated with  
18 the Columbia Plateau, or it could be the result of spread from separate sets of refugia after the last glacial  
19 maximum.

## 20 Keywords

21 *Clarkia*, isolation-by-distance, landscape genetics, population genetics, genetic diversity

## 22 Introduction

23 Geographic distance is often a primary predictor of genetic differentiation among populations on the land-  
24 scape. Populations that are near each other are often more genetically similar, while distant populations are  
25 often more divergent. This pattern arises when the dispersal distances of individuals and gametes are small  
26 relative to the distances separating populations; as a result, differences accumulate among populations due  
27 to drift faster than they are homogenized by gene flow (Slatkin, 1993; Wright, 1943). Isolation by distance  
28 is well-documented and prevalent (Sexton et al., 2014) to the extent that it is a reasonable null expectation  
29 for how genetic differentiation is structured at geographic scales.

30 However, geographic distance is not the only factor that structures dispersal and realized gene flow among  
31 populations (McRae, 2006; Epps et al., 2005). Not all geographic distances are equivalent in the extent to  
32 which they might facilitate or impede gene flow (Storfer et al., 2007). Landscape features between populations  
33 may impose barriers to gene flow beyond those predicted by geographic distance. Gaps in suitable habitat  
34 may be large enough that very few instances of gene flow occur across them, leading to differentiation of the  
35 populations on either side. For example, Reeves and Richards (2014) found genetic differentiation between  
36 populations of *Helianthus pumilus* that could be attributed to an unsuitable mountainous area interrupting  
37 the species' distribution. Other features of the landscape might act as corridors for the organisms themselves  
38 or for agents of gene flow (i.e. seed dispersers or pollinators). For example, wind and water flow along rivers  
39 may increase gene flow among populations situated along them (Lee et al., 2018). In these types of scenarios  
40 we expect to see deviations from a strict pattern of isolation by distance, and population genetic structure  
41 will be better described by membership in discrete groups on either side of a barrier in the former case, or  
42 by patterns of admixture or increased similarity in populations connected by corridors in the latter.

43 Environmental differences between occupied sites may also contribute to the magnitude of genetic differ-  
44 entiation between populations (Slatkin, 1973; Wang and Bradburd, 2014). If populations are strongly locally  
45 adapted, then migrants that have moved between environments may be unable to survive to reproduction or  
46 may have low reproductive success (Nosil et al., 2005). In this case, realized gene flow may be low between  
47 different environments (Mosca et al., 2012). Similarly, vectors of gene flow such as pollinators and seed  
48 dispersers (or the organisms themselves, in the case of motile species) may have environmental preferences  
49 that lead to greater rates of gene flow among similar environments (Bolnick et al., 2009).

50 The current genetic structure of populations is also strongly influenced by past processes (Hewitt, 2004).  
51 In temperate regions including the Pacific Northwest, higher latitudes were glaciated until approximately  
52 20,000 years ago (Booth et al., 2003) and this affected the distribution of many species, leaving lasting  
53 signatures on their genetic structure (Brunsfeld et al., 2001; Shafer et al., 2010). Species that previously  
54 had disjunct distributions—for example, those that occupied multiple refugia during glaciation—may exhibit  
55 multiple corresponding genetic clusters in the present day (Beatty and Provan, 2011; Carstens et al., 2013;  
56 Sproul et al., 2015). Populations that are the result of range expansions into previously glaciated areas  
57 may have lower levels of genetic diversity as a result of repeated founder events (Kuchta and Tan, 2005;  
58 Hewitt, 2004). These patterns may underlie (and sometimes confound) genetic structure that could also be  
59 attributed to isolation by distance or environment.

60 Despite the accumulation of numerous case studies, it is still challenging to draw generalizations about  
61 the extent to which the genetic structure of a given species is likely to be determined by geographic vs.  
62 environmental differences. A recent meta-analysis (Sexton et al., 2014) examined how the frequency of

63 isolation by distance vs. by environment varied across broad taxonomic groups, and found that plants more  
64 frequently showed patterns of isolation by distance than vertebrates or invertebrates. However, in more  
65 than half of the plant species that displayed a pattern of isolation by distance, environmental similarity  
66 also contributed to genetic structure. In a small number of plant species, only environmental differences  
67 explained genetic structure. Although geography and environment may both have important effects on  
68 patterns of genetic differentiation, generalizations about when one will prevail over the other and what  
69 organismal traits determine their relative effect sizes remain elusive. The accumulation of more case studies  
70 and the development and use of more appropriate statistical methods will likely move this field forward  
71 (Wang and Bradburd, 2014; Bradburd et al., 2013).

72 The way that the landscape shapes genetic structure is of particular interest in the context of geographic  
73 range limits. Local adaptation may be constrained in range edge populations if these populations are  
74 inundated with gene flow from populations in dissimilar environments (Kirkpatrick and Barton, 1997). If  
75 populations are isolated by environmental differences, that might prevent swamping gene flow. Rather, gene  
76 flow between populations in similar environments could facilitate local adaptation by increasing adaptive  
77 genetic diversity (Sexton et al., 2011). This might be of particular importance if species occupy spatially  
78 heterogeneous environments, where random dispersal would otherwise result in frequent gene flow between  
79 divergent environments.

80 In this study, we investigated whether environmental differences between populations of the annual  
81 wildflower *Clarkia pulchella* contribute to their genetic differentiation, which we expected to also be strongly  
82 structured by geographic distances. Further, we explored whether patterns of genetic differentiation are  
83 better described by admixture among distinct genetic groups or continuous genetic differentiation across the  
84 landscape. We expected that topographic features, such as the Rocky Mountains, might be an impediment  
85 to the movement of seed dispersers and pollinators, and that this might result in disjunct genetic groups.  
86 Finally, we explored whether genetic diversity varies geographically in this species. We predicted lower levels  
87 of genetic diversity at high latitudes if this species has undergone a range expansion northward after the last  
88 glacial maximum.

## 89 Methods

### 90 Study species

91 *Clarkia pulchella* Pursh (Onagraceae) is a winter annual wildflower that grows east of the Cascade Mountains  
92 in the Pacific Northwest. It can be found in eastern Washington, eastern Oregon, Idaho, and western

93 Montana (United States) and in southeastern British Columbia (Canada; Figure 1). It grows in large  
94 populations (i.e., thousands of flowering individuals) on open, south-facing slopes from 100 to 2200 meters  
95 elevation, though the majority of populations are found between 500 and 1600 m. While temperature  
96 generally decreases and precipitation generally increases from south to north and west to east across the  
97 range of *C. pulchella*, temperature and precipitation are also strongly influenced by elevation. Topographic  
98 complexity across the range creates large amounts of variation around geographic trends and appears to  
99 disrupt spatial autocorrelation in climate among populations of *C. pulchella* (Figure 2). This species has  
100 small seeds (c. 1 mm long) that lack an obvious dispersal mechanism. Flowers are visited by a diverse array  
101 of pollinators, including solitary bees, bee flies, bumblebees, and occasionally hummingbirds (M. Bontrager,  
102 personal observation).

### 103 **Population selection, climate characterization, and seed collection**

104 For this study, we selected populations that would allow us to decouple climatic and spatial axes of differen-  
105 tiation. For example, we wanted to include populations that were spatially near each other but climatically  
106 different and populations that were geographically distant but climatically similar. Monthly temperature and  
107 precipitation data from 1951-1980 for all populations were obtained from PRISM (PRISM Climate Group,  
108 2017). We calculated the average temperature across the months that encompass the *C. pulchella* life cycle  
109 (September-July) and average precipitation when *C. pulchella* is most likely to be water-limited (April-July)  
110 for each population. Based on field observations and common garden trials (Bontrager and Angert, in prep),  
111 we considered these to be good candidates for variables that might have the potential to generate patterns  
112 of isolation by environment via selection against migrants. We first considered a set of 40 populations that  
113 we had located, then narrowed that set down to 32 populations that maximized variation in the relationship  
114 between spatial proximity and climatic similarity (Figure 1, Table S1). In July of 2014, we collected seeds  
115 from 12 plants separated by at least 0.5 m in each of those populations. Seeds from 17 populations were  
116 grown in the greenhouse beginning in December of 2014, and seeds from the remaining 15 populations were  
117 grown in growth chambers beginning in February of 2016.

### 118 **DNA Extraction**

119 Tissue was harvested from the first cohort of plants in May 2015. Leaf or bud tissue was collected into 2 mL  
120 tubes on ice, then frozen at -80°C until DNA extraction. Tissue from the second cohort was collected onto  
121 dry ice in April 2016 and stored at -80°C until DNA extraction. DNA was extracted using DNeasy Plant  
122 Mini kits and DNeasy Plant 96 kits (Qiagen), following the protocol for frozen tissues. DNA extractions

123 that did not have satisfactory 260/230 or 260/280 ratios were cleaned with ethanol precipitation. DNA was  
124 eluted and stored in 10mM Tris-HCl pH 8.

125 **Library preparation and sequencing**

126 Libraries were prepared using 100 ng starting material. We prepared for two lanes of sequencing, with six  
127 individually barcoded samples from each population in each lane (191 or 192 individuals per lane, because  
128 we only had DNA of a high enough quality from a total 11 individuals from one population). Our library  
129 preparation protocol was based on Poland et al. (2012) with modification by M. Todesco, K. Ostevik, and  
130 B. Moyers (Rieseberg Lab, University of British Columbia). DNA was digested in a 20  $\mu$ L reaction using 8  
131 units each of the enzymes MspI and Pst I-HF (New England Biolabs) in the supplied buffer. Digestion was  
132 carried out for 5 hours at 37°C, followed by 20 minutes at 65°C. Reactions were then stored overnight at  
133 4°C. Ligation was performed in a 40  $\mu$ L reaction in the same buffer as the digestion with 200 units of T4  
134 DNA ligase (New England Biolabs) using 192 barcoded adapters and 12 common adapters on the opposite  
135 end. Ligation was performed for 3 hours at 22°C followed by a 20 minute hold at 65°C. Reactions were then  
136 cleaned with 1.6 volumes of SPRI beads and two 80% ethanol washes and resuspended in 12  $\mu$ L of Tris-HCl  
137 pH 8.

138 Amplification was carried out in 10  $\mu$ L reactions using 4  $\mu$ L of cleaned ligation product, Kapa HiFi  
139 HotStart master mix (Kapa Biosystems), and primers from Poland et al. (2012). Amplification began at  
140 98°C (30 s), followed by 14 cycles of 98°C (30 s), 62°C (20 s), 72°C (30 s), and a 72°C hold for 5 minutes.  
141 After amplification, samples were quantified using fluorometry, then each plate was pooled according to  
142 individual concentrations to yield a final product with equal amounts of library from each individual. This  
143 pooled library was run out on a 1.5% agarose gel and bands containing fragments 400 to 600 bp long were  
144 excised and cleaned using a gel extraction kit (Qiagen). The eluted product was cleaned and concentrated  
145 using SPRI beads.

146 Finally, we reduced the number of high copy fragments from our library using a protocol modified by M.  
147 Todesco from Shagina et al. (2010) and Matvienko et al. (2013). We began with 480 ng of each library in  
148 a 3  $\mu$ L volume. To this we added 1  $\mu$ L of hybridization buffer (200 mM HEPES pH 7.5, 2M NaCl, 0.8 mM  
149 EDTA), covered the reaction with mineral oil, heated it to 98°C for 2 minutes, then held it at 78°C for 3  
150 hours. We then added 5  $\mu$ L of duplex specific nuclease buffer (0.1 M Tris pH 8, 10mM MgCl<sub>2</sub>, 2mM DTT)  
151 and incubated at 70°C for 5 minutes. We then added 0.2  $\mu$ L of duplex specific nuclease and incubated at  
152 70°C for another 15 minutes, then stopped the reaction with 10  $\mu$ L of 10 mM EDTA. We then reamplified  
153 the library using the same reagents as above in a 25  $\mu$ L reaction with 2-4  $\mu$ L of template and cleaned again

154 with SPRI beads. Libraries were stored at -20°C until sequencing. Libraries were sequenced with paired-end  
155 100 bp reads on the Illumina HiSeq 2000 platform at the Biodiversity Research Centre at UBC.

## 156 Alignment and SNP calling

157 Sequences were processed and aligned using components of the Stacks pipeline (version 1.40, Catchen et al.,  
158 2011, 2013). Reads with uncalled bases or low quality scores (average quality in a 14-base sliding window  
159 <10) were discarded. Ten samples had far fewer reads than the rest and these were excluded prior to  
160 alignment. Paired end reads were pooled with first end reads, i.e. during alignment and SNP (single  
161 nucleotide polymorphism) detection the two ends of each read were treated as if they were independent loci  
162 (we later checked for linkage disequilibrium among SNPs). During initial “stacking” and catalog building we  
163 allowed sequences to diverge at 3 bases, and set the minimum depth of coverage required to create a stack at  
164 3 (Rochette and Catchen, 2017). Modifications to these parameters did not result in substantial differences  
165 in values of pairwise  $F_{ST}$  (data not shown). The maximum number of stacks per locus was set to 3, and  
166 gapped alignments were not allowed. We enabled the removal algorithm, which drops highly repetitive stacks  
167 (removes initial stacks that have >2 SD coverage relative to individual sample mean), and the deleveraging  
168 algorithm, which breaks up or removes over-merged sequences. Our catalog was built using all samples. We  
169 employed the rxstacks corrections module to correct or omit loci with putative sequencing errors, loci with  
170 low log-likelihoods (<-10), confounded loci, and loci with excess haplotypes.

171 SNP tables were generated using the populations module of Stacks. Initial inspection of PCA plots using  
172 SNPRelate (Zheng et al., 2012) revealed three individuals that were not clustering with the other individuals  
173 from their populations. We consider it more plausible that these represent mis-labeled samples in the field,  
174 greenhouse, or lab than long-distance migration events. Downstream analyses were performed without these  
175 individuals. Therefore, in our final dataset, seven populations had only 11 individuals, one population  
176 had only 10, one population had only 8, and the remaining 23 populations were each represented by 12  
177 individuals. In our analyses we included only loci that had coverage of at least 12x in 75% of individuals  
178 in 75% of populations, with a minimum minor allele frequency of 0.05 and a maximum heterozygosity of  
179 70% across all populations. We checked that pairwise  $F_{ST}$  was not sensitive to these parameter choices. In  
180 case of multiple SNPs occurring in a single locus, we kept just the first one. After applying these filters,  
181 2983 SNPs were retained. Linkage disequilibrium was generally low among our loci ( $r^2 < 0.2$  for 99.9% of  
182 pairs of SNPs).  $F_{ST}$  was calculated using the implementation of Weir and Cockerham (1984) and expected  
183 heterozygosity (within-population gene diversity) was calculated using methods from Nei (1987) in the R  
184 package hierfstat (Goudet and Jombart, 2015). Because populations varied in the average proportion of loci

185 that were successfully genotyped (three populations had <60% success; among all populations the median  
186 success rate was 78% and the range was 23-92%), we checked that expected heterozygosity did not correlate  
187 with genotyping success rate ( $r = 0.27$ ,  $P = 0.13$ ).

## 188 Quantifying isolation by environment vs. isolation by distance

189 We used BEDASSLE (Bradburd et al., 2013) to estimate the relative contributions of geographic distance  
190 and climatic differences to genetic differentiation. BEDASSLE is implemented in R (R Core Team, 2017),  
191 and it employs a Markov chain Monte Carlo (MCMC) algorithm to estimate the relative effect sizes of  
192 geographic distance and environmental differences on covariance in allele frequencies among populations. As  
193 environmental covariates, we used pairwise differences in average September-July temperature and average  
194 spring/summer precipitation (April-July). We initially generated resistance-weighted distances between  
195 populations using projected habitat suitability (Chapter 1) as a conductance matrix, but these distances  
196 were highly correlated with actual geographic distances and did not produce better model fits in preliminary  
197 analyses, so we did not use them in these models. We estimated effect sizes of geography, temperature,  
198 and precipitation differences using all 32 populations, but also ran BEDASSLE for subsets consisting of  
199 populations clustered in the central and northern parts of the range (indicated in Table S1) to see if we  
200 could detect effects of the environment that may be obscured or weakened at large geographic scales. Prior  
201 to analysis, we divided pairwise geographic distance and precipitation differences by their standard deviations  
202 so that these predictors were on a scale more similar to pairwise temperature differences. We ran these models  
203 for 10 million generations, and thinned the chains by sampling every 1000 generations. We visually inspected  
204 MCMC traces and marginal distributions to ensure that models reached stationary distributions. All results  
205 are reported after a burn-in of 20%, with effect sizes back-transformed to the scale of the original data.  
206 We checked these results against partial Mantel tests of pairwise geographic, temperature, and precipitation  
207 differences on pairwise  $F_{ST}$  using the R package phytools (Revell, 2012). We did not rely upon partial Mantel  
208 tests as our main analytical method because of their potential to have inflated Type I error rates (Guillot  
209 and Rousset, 2013).

## 210 Assessment of spatially continuous vs. discrete genetic differentiation

211 We were interested in evaluating whether population structure was well-described by modelling populations  
212 as admixtures between multiple discrete genetic groups, as might be caused by geographic barriers (i.e., the  
213 Rocky Mountains) or historic phylogeographic processes. We evaluated how well models prescribing various  
214 numbers of discrete genetic groups described differentiation and similarity among our populations using

215 conStruct (Bradburd et al., 2017). conStruct is implemented in R (R Core Team, 2017), and is similar to the  
216 frequently-used program Structure (Pritchard et al., 2000) but allows genetic differentiation to increase with  
217 geographic distance between populations even when these populations draw from the same genetic groups.  
218 In the spatial implementation of this program, populations are composed of admixture from a user-specified  
219 number of discrete layers (K), and genetic similarity decays with geographic distance within each of these  
220 layers. We ran conStruct for 1000 iterations setting the number of layers to 1, 2, 3, 4, and 5. We compared  
221 the fits of each of these different parameterizations using cross-validation and by evaluating the contribution  
222 of each additional layer to the total covariance of these loci. For cross-validation, we fit models with subsets  
223 containing 90% of loci and evaluated the resulting model fit by calculating the log likelihood of the remaining  
224 loci. We performed 100 replicate cross-validation runs.

## 225 Exploring spatial patterns in genetic diversity

226 We examined whether population genetic diversity (as estimated by expected heterozygosity) exhibited  
227 geographic trends. We used linear models in R (R Core Team, 2017) to test whether expected heterozygosity  
228 was predicted by latitude or by proximity to the range edge (as measured by the distance of a population  
229 to the nearest edge of a polygon drawn around all localities of the species; Figure 1).

## 230 Results

### 231 Isolation by environment vs. geographic distance

232 Overall  $F_{ST}$  among these populations is 0.135. Genetic differentiation between populations of *Clarkia pul-*  
233 *chella* is primarily structured by geographic distance, with no apparent contribution of the environmental  
234 variables that we have considered here (Figure 3). The effect size of a temperature difference of one degree  
235 (C) relative to the effect of 100 km of geographic distance is  $1.18 \times 10^{-7}$  (95% credible interval =  $8.52 \times 10^{-8}$   
236 -  $1.58 \times 10^{-7}$ ; Figure S1A), and the effect of 10 mm of spring/summer precipitation difference relative to  
237 the effect of 100 km of geographic distance is  $5.84 \times 10^{-7}$  (95% credible interval =  $1.50 \times 10^{-8}$  -  $2.98 \times 10^{-6}$ ;  
238 Figure S1B). The scales at which these ratios are presented are arbitrary, but they were chosen so that the  
239 range of values among populations is on the same order of magnitude: 100 km represents about one sixth of  
240 the maximum pairwise geographic distance, 1°C represents approximately one fourth of the maximum pair-  
241 wise temperature difference, and 10 mm precipitation represents about one fourth of the maximum pairwise  
242 precipitation difference. The climatic effect sizes we found are so small that the effects of these variables  
243 can be considered nonexistent in terms of their biological importance. Effects of environmental differences

244 did not emerge at smaller geographic scales in subsets of populations in the north (effect of temperature  
245 differences relative to geographic distance:  $5.89 \times 10^{-8}$  ( $8.61 \times 10^{-9}$  -  $1.14 \times 10^{-7}$ ), effect of precipitation  
246 differences relative to geographic distance:  $9.73 \times 10^{-6}$  ( $5.81 \times 10^{-7}$  -  $2.11 \times 10^{-5}$ ); Figure S2) or center (effect  
247 of temperature:  $2.34 \times 10^{-7}$  ( $1.44 \times 10^{-8}$  -  $4.73 \times 10^{-7}$ ), effect of precipitation:  $9.46 \times 10^{-7}$  ( $3.06 \times 10^{-8}$  -  $4.80$   
248  $\times 10^{-6}$ ); Figure S3). These conclusions are consistent with the results of partial Mantel tests, in which only  
249 pairwise geographic distance is a significant predictor of pairwise  $F_{ST}$  (Table 1).

## 250 **Genetic structure of populations**

251 The genetic structure of these populations is explained slightly better by a model of admixture between  
252 two genetic groups than by a model of continuous genetic differentiation across space, as indicated by  
253 the increase in predictive accuracy in models where two layers were allowed rather than one (Figure 4).  
254 Northern populations primarily belong to one genetic group, while southern populations belong to another,  
255 and populations from mid-latitudes are a mix of the two (Figure 5). Allowing more than two layers did not  
256 improve predictive accuracy (Figure 4). Note that populations east of the Rocky Mountains (populations  
257 D9, D10, and P12) never formed a separate group, regardless of the number of layers allowed (results not  
258 shown). Although models with two layers did have greater predictive accuracy than those with one, when  
259  $K = 2$  the amount of covariance contributed by the second layer was small relative to the first (Table 2).

## 260 **Geographic trends in genetic diversity**

261 Genetic diversity increases with latitude among these populations (estimate = 0.0104, SE = 0.0019, df = 30,  
262  $P < 0.0001$ , Figure 6A), but is not related to distance from the range edge (df = 30,  $P = 0.811$ ). Genetic  
263 diversity appears to be lower in populations in the southern half of the range, and also in populations near  
264 the eastern range edge, but is higher in central and northern populations (Figure 6B).

## 265 **Discussion**

266 We contrasted the relative effects of geographic vs. climatic distances on genetic differentiation in *Clarkia*  
267 *pulchella*, examined whether geographic structure in this species could be described by assigning populations  
268 to distinct genetic groups, and tested for geographic gradients in genetic diversity. Our analyses revealed a  
269 genetic structure that is predominantly shaped by geographic distances between populations. In addition to  
270 this pattern of isolation by distance, populations partition into northern and southern groups, with admixed  
271 populations in the center of the range. Genetic diversity was highest in northern and central populations,  
272 resulting in a trend of increasing genetic diversity with latitude.

273 **Populations of *Clarkia pulchella* are isolated by distance**

274 At the scale of the geographic range in *Clarkia pulchella*, isolation by distance is the dominant pattern.  
275 This likely reflects gene flow that is strongly restricted by geographic distances between populations. This  
276 is perhaps not surprising, given that this species has no obvious mechanism for seed dispersal and our best  
277 guess is that gene flow between populations is facilitated by occasional pollen movement by bumblebees,  
278 hummingbirds, and other floral visitors. In the case of an absence of climatically structured seed and  
279 pollen movement, selection against migrants and their offspring is the remaining mechanism that could drive  
280 isolation by environment. While *C. pulchella* does appear to be locally adapted to historic climate (Bontrager  
281 and Angert, in prep), selection against foreign genotypes may not be strong enough to preempt the spread  
282 of neutral loci, even as recently-arrived loci that confer poor performance in a given environment are purged.  
283 This could lead to a signal of isolation by distance at neutral loci, while populations are still adaptively  
284 differentiated based on their local climate.

285 It is possible that the absence of an effect of temperature and precipitation differences on genetic structure  
286 is the result of our experimental design, and that environmental differences might matter in other contexts.  
287 There may be environmental variables other than those we have considered here that are more important  
288 in determining the movement of genes or the realized rate of gene flow among populations. These could be  
289 climatic, but also could include soil characteristics, or local adaptation to competitors, pollinators, or soil  
290 biota. It is also possible that the effects of environmental differences are more detectable at smaller spatial  
291 scales. For example, in some plant species, differences in phenological timing along local snowmelt gradients  
292 structure gene flow to a greater extent than geographic distances (Hirao and Kudo, 2004; Shimono et al.,  
293 2009). Similar processes may play out in *C. pulchella* as well, possibly along local elevation gradients.

294 **Populations are admixtures of northern and southern genetic groups**

295 Rather than mountain ranges separating populations into genetic groups, we detected underlying population  
296 structure that divides the species into northern and southern groups, with admixed populations in the  
297 middle. This suggests that perhaps the Columbia Basin, a low-elevation, relatively flat area in south-central  
298 Washington (Figure 1), is a barrier to gene flow in this species. Species distribution models indicate that it  
299 is an area of low suitability (Chapter 1) and few occurrences of *Clarkia pulchella* have been recorded in this  
300 region. Most studies of population genetic structure in the Pacific Northwest focus on mesic forest species  
301 that occupy the wet western slopes of both the coastal and Rocky Mountains (Shafer et al., 2010), and these  
302 studies often find differentiation between western and eastern populations. Phylogeographic research on  
303 species occupying the arid inter-mountain region is less common. In the Great Basin pocketmouse, a species

304 with a range that overlaps with that of *C. pulchella*, a north-south split in genetic structure was detected  
305 in approximately the same location as in our results (Riddle et al., 2014). It is possible that the Columbia  
306 Basin (or some geographic feature within it) represents a barrier to gene flow, either past or ongoing, for a  
307 variety of taxa that occupy the dry intermountain region. The habitat affinity of species can influence the  
308 effect of glaciation events on genetic structure (Massatti and Knowles, 2014), therefore further work on *C.*  
309 *pulchella*, including paleoclimate modelling or modelling demographic history, might allow for an interesting  
310 contrast with the relatively well-studied mesic flora of the Pacific Northwest.

### 311 **Genetic diversity increases with latitude**

312 We expected we would see lower genetic diversity at higher latitudes, but we detected the opposite: genetic  
313 diversity was highest in north-central and northern populations (though the total magnitude of variation in  
314 expected heterozygosity was not large). This latitudinal pattern is somewhat surprising, because northern  
315 populations are in areas that were under glaciers during the last glacial maximum, and we expected that  
316 range expansion into this area after their retreat would result in a signature of lower genetic diversity.  
317 When high levels of genetic diversity are present in areas of past range expansion, this can sometimes be  
318 attributed to the mixing of populations that had previously been persisting in multiple refugia (Petit et al.,  
319 2003; Brunsfeld and Sullivan, 2005). Species in the northern Rocky Mountains that are presumed to have  
320 occupied multiple refugia often exhibit some degree of contemporary differentiation between northern and  
321 southern populations (Brunsfeld et al., 2001; Brunsfeld and Sullivan, 2005), a pattern consistent with what we  
322 have found in *Clarkia pulchella*. Regardless of the location or number of refugia that *C. pulchella* previously  
323 occupied, it is also possible that range expansion was not accompanied by reductions in genetic diversity in  
324 this species, as is sometimes the case in other systems (Vandepitte et al., 2017).

325 The more common expectation for geographic patterns in genetic diversity is that range edge populations  
326 will have lower genetic diversity (Vucetich and Waite, 2003). This prediction is based on the assumption of  
327 an abundant center distribution pattern, in which edge populations are small, and may experience frequent  
328 turnover or constant directional selection (if they are far from the phenotypic optima of an extreme environ-  
329 ment). Our results are not consistent with this being the case for *C. pulchella*, at least not at all range edges.  
330 We note however that populations at southern and eastern edges do appear to have lower genetic diversity  
331 relative to the northern and north-central populations, and further work could be done to investigate the  
332 processes that might generate this pattern.

### 333 **Conclusions**

334 Our investigation of the genetic structure of *Clarkia pulchella* has revealed some intuitive patterns, as well as  
335 surprising ones. Despite substantial heterogeneity in climate across the species' range, genetic similarity is  
336 primarily determined by geographic proximity. Though a signal of isolation by distance is not surprising in a  
337 sessile organism studied at a large spatial scale, the absence of any effect of environment indicates that to the  
338 extent that populations experience gene flow, it may be from both similar and divergent environments. This  
339 species does not exhibit geographic patterns of genetic diversity consistent with our expectations for a recently  
340 expanded northern range edge nor a range limited by adaptation. These results would be complemented by  
341 future work examining mechanisms of contemporary gene flow and historic demographic processes in *Clarkia*  
342 *pulchella*.

### 343 **Acknowledgements**

344 We would like to thank C. Caseys and M. Todesco for their generous guidance and training during library  
345 preparation. G. Owens provided helpful advice on bioinformatics methods. E. Fitz assisted with locating  
346 populations of *C. pulchella* in the field and A. Wilkinson assisted with plant cultivation. We appreciate  
347 thoughtful comments on the manuscript from S. Aitken, M. Whitlock, and J. Whitton. Funding for this  
348 project was provided by the Washington Native Plant Society, the Botanical Society of America and the  
349 Botanical Society of America Genetics Section, and a National Sciences and Engineering Research Council  
350 Discovery Grant to ALA. MB was supported by a University of British Columbia Four-year Fellowship.

### 351 **Author contributions**

352 MB conceived of the project in consultation with ALA. MB performed all field work, lab work, bioinformatics,  
353 and analyses with guidance from ALA. MB generated all figures and tables and wrote the manuscript with  
354 frequent conversation and comments from ALA.

### 355 **Data accessibility**

356 Data and code are hosted on Github at <https://github.com/meganbontrager/clarkia-pulchella-popgen> and  
357 will be archived on Dryad or a similar repository upon publication. Sequences are archived on the NCBI  
358 Sequence Read Archive (SUB4307183).

359 **References**

- 360 Beatty, G. E. and J. Provan (2011). Phylogeographic analysis of North American populations of the parasitic  
361 herbaceous plant *Monotropa hypopitys* L. reveals a complex history of range expansion from multiple late  
362 glacial refugia. *Journal of Biogeography* 38(8), 1585–1599.
- 363 Bolnick, D. I., L. K. Snowberg, C. Patenia, W. E. Stutz, T. Ingram, and O. L. Lau (2009). Phenotype-  
364 dependent native habitat preference facilitates divergence between parapatric lake and stream stickleback.  
365 *Evolution* 63(8), 2004–2016.
- 366 Booth, D. B., K. G. Troost, J. J. Clague, and R. B. Waitt (2003). The Cordilleran ice sheet. *Developments*  
367 *in Quaternary Sciences* 1, 17–43.
- 368 Bradburd, G., G. Coop, and P. Ralph (2017). Inferring continuous and discrete population genetic structure  
369 across space. *bioRxiv*, 189688.
- 370 Bradburd, G. S., P. L. Ralph, and G. M. Coop (2013). Disentangling the effects of geographic and ecological  
371 isolation on genetic differentiation. *Evolution* 67(11), 3258–3273.
- 372 Brunsfeld, S., J. Sullivan, D. Soltis, and P. Soltis (2001). Comparative phylogeography of northwestern  
373 North America: a synthesis. *Special Publication–British Ecological Society* 14, 319–340.
- 374 Brunsfeld, S. J. and J. Sullivan (2005). A multi-compartmented glacial refugium in the northern Rocky  
375 Mountains: evidence from the phylogeography of *Cardamine constancei* (Brassicaceae). *Conservation*  
376 *Genetics* 6(6), 895–904.
- 377 Carstens, B. C., R. S. Brennan, V. Chua, C. V. Duffie, M. G. Harvey, R. A. Koch, C. D. McMahan, B. J.  
378 Nelson, C. E. Newman, J. D. Satler, et al. (2013). Model selection as a tool for phylogeographic inference:  
379 an example from the willow *Salix melanopsis*. *Molecular Ecology* 22(15), 4014–4028.
- 380 Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko (2013). Stacks: an analysis tool  
381 set for population genomics. *Molecular Ecology* 22(11), 3124–3140.
- 382 Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait (2011). Stacks: building and  
383 genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* 1(3), 171–182.
- 384 Epps, C. W., P. J. Palsbøll, J. D. Wehausen, G. K. Roderick, R. R. Ramey, and D. R. McCullough (2005).  
385 Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. *Ecology*  
386 *Letters* 8(10), 1029–1038.

- 387 Goudet, J. and T. Jombart (2015). hierfstat: Estimation and tests of hierarchical F-statistics. R package  
388 version 0.04-22.
- 389 Guillot, G. and F. Rousset (2013). Dismantling the Mantel tests. *Methods in Ecology and Evolution* 4(4),  
390 336–344.
- 391 Hewitt, G. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions  
392 of the Royal Society B: Biological Sciences* 359(1442), 183–195.
- 393 Hirao, A. and G. Kudo (2004). Landscape genetics of alpine-snowbed plants: comparisons along geographic  
394 and snowmelt gradients. *Heredity* 93(3), 290.
- 395 Kirkpatrick, M. and N. H. Barton (1997). Evolution of a species' range. *The American Naturalist* 150(1),  
396 1–23.
- 397 Kuchta, S. R. and A.-M. Tan (2005). Isolation by distance and post-glacial range expansion in the rough-  
398 skinned newt, *Taricha granulosa*. *Molecular Ecology* 14(1), 225–244.
- 399 Lee, S.-R., Y.-S. Jo, C.-H. Park, J. M. Friedman, and M. S. Olson (2018). Population genomic analysis  
400 suggests strong influence of river network on spatial distribution of genetic variation in invasive saltcedar  
401 across the southwestern United States. *Molecular Ecology* 27(3), 636–646.
- 402 Massatti, R. and L. L. Knowles (2014). Microhabitat differences impact phylogeographic concordance of  
403 codistributed species: Genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evo-  
404 lution* 68(10), 2833–2846.
- 405 Matvienko, M., A. Kozik, L. Froenicke, D. Lavelle, B. Martineau, B. Perroud, and R. Michelmore (2013).  
406 Consequences of normalizing transcriptomic and genomic libraries of plant genomes using a duplex-specific  
407 nuclease and tetramethylammonium chloride. *PLoS One* 8(2), e55913.
- 408 McRae, B. H. (2006). Isolation by resistance. *Evolution* 60(8), 1551–1561.
- 409 Mosca, E., A. Eckert, E. Di Pierro, D. Rocchini, N. La Porta, P. Belletti, and D. Neale (2012). The  
410 geographical and environmental determinants of genetic diversity for four alpine conifers of the European  
411 Alps. *Molecular Ecology* 21(22), 5530–5545.
- 412 Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press.
- 413 Nosil, P., T. H. Vines, and D. J. Funk (2005). Perspective: reproductive isolation caused by natural selection  
414 against immigrants from divergent habitats. *Evolution* 59(4), 705–719.

- 415 Petit, R. J., I. Aguinagalde, J.-L. de Beaulieu, C. Bittkau, S. Brewer, R. Cheddadi, R. Ennos, S. Fineschi,  
416 D. Grivet, M. Lascoux, et al. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity.  
417 *Science* 300(5625), 1563–1565.
- 418 Poland, J. A., P. J. Brown, M. E. Sorrells, and J.-L. Jannink (2012). Development of high-density genetic  
419 maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS One* 7(2),  
420 e32253.
- 421 PRISM Climate Group (2017). Oregon state university. Accessed: 2017-09-30.
- 422 Pritchard, J. K., M. Stephens, and P. Donnelly (2000). Inference of population structure using multilocus  
423 genotype data. *Genetics* 155(2), 945–959.
- 424 R Core Team (2017). *R: a language and environment for statistical computing, version 3.4.4*. Vienna,  
425 Austria: R Foundation for Statistical Computing.
- 426 Reeves, P. A. and C. M. Richards (2014). Effect of a geographic barrier on adaptation in the dwarf sunflower  
427 (*Helianthus pumilus* Nutt.). *International Journal of Plant Sciences* 175(6), 688–701.
- 428 Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things).  
429 *Methods in Ecology and Evolution* 3, 217–223.
- 430 Riddle, B. R., T. Jezkova, M. E. Eckstut, V. Oláh-Hemmings, and L. N. Carraway (2014). Cryptic divergence  
431 and revised species taxonomy within the Great Basin pocket mouse, *Perognathus parvus* (Peale, 1848),  
432 species group. *Journal of Mammalogy* 95(1), 9–25.
- 433 Rochette, N. C. and J. M. Catchen (2017). Deriving genotypes from RAD-seq short-read data using Stacks.  
434 *Nature Protocols* 12(12), 2640.
- 435 Sexton, J. P., S. B. Hangartner, and A. A. Hoffmann (2014). Genetic isolation by environment or distance:  
436 which pattern of gene flow is most common? *Evolution* 68(1), 1–15.
- 437 Sexton, J. P., S. Y. Strauss, and K. J. Rice (2011). Gene flow increases fitness at the warm edge of a species  
438 range. *Proceedings of the National Academy of Sciences* 108(28), 11704–11709.
- 439 Shafer, A., C. I. Cullingham, S. D. Cote, and D. W. Coltman (2010). Of glaciers and refugia: a decade of  
440 study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology* 19(21),  
441 4589–4621.
- 442 Shagina, I., E. Bogdanova, I. Z. Mamedov, Y. Lebedev, S. Lukyanov, D. Shagin, et al. (2010). Normalization  
443 of genomic DNA using duplex-specific nuclease. *Biotechniques* 48(6), 455.

- 444 Shimono, Y., M. Watanabe, A. S. Hirao, N. Wada, and G. Kudo (2009). Morphological and genetic variations  
445 of *Potentilla matsumurae* (Rosaceae) between fellfield and snowbed populations. *American Journal of*  
446 *Botany* 96(4), 728–737.
- 447 Slatkin, M. (1973). Gene flow and selection in a cline. *Genetics* 75(4), 733–756.
- 448 Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47(1),  
449 264–279.
- 450 Sproul, J. S., D. D. Houston, C. R. Nelson, R. P. Evans, K. A. Crandall, and D. K. Shiozawa (2015). Climate  
451 oscillations, glacial refugia, and dispersal ability: factors influencing the genetic structure of the least  
452 salmonfly, *Pteronarcella badia* (Plecoptera), in western North America. *BMC Evolutionary Biology* 15(1),  
453 279.
- 454 Storfer, A., M. Murphy, J. Evans, C. Goldberg, S. Robinson, S. Spear, R. Dezzani, E. Delmelle, L. Vierling,  
455 and L. Waits (2007). Putting the ‘landscape’ in landscape genetics. *Heredity* 98(3), 128.
- 456 Vandepitte, K., K. Helsen, K. Van Acker, J. Mergeay, and O. Honnay (2017). Retention of gene diversity  
457 during the spread of a non-native plant species. *Molecular Ecology* 26(12), 3141–3150.
- 458 Vucetich, J. A. and T. A. Waite (2003). Spatial patterns of demography and genetic processes across the  
459 species’ range: null hypotheses for landscape conservation genetics. *Conservation Genetics* 4(5), 639–645.
- 460 Wang, I. J. and G. S. Bradburd (2014). Isolation by environment. *Molecular Ecology* 23(23), 5649–5662.
- 461 Weir, B. S. and C. C. Cockerham (1984). Estimating F-statistics for the analysis of population structure.  
462 *Evolution* 38(6), 1358–1370.
- 463 Wright, S. (1943). Isolation by distance. *Genetics* 28(2), 114–138.
- 464 Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir (2012). A high-performance  
465 computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28(24),  
466 3326–3328.

Table 1 Results of partial Mantel tests of pairwise geographic distance (km), pairwise temperature differences ( $^{\circ}\text{C}$ , September-July, 1951-1980 averages), and pairwise precipitation differences (mm, April-July, 1951-1980 averages) on pairwise genetic differentiation ( $F_{\text{ST}}$ ) among populations of *Clarkia pulchella*. Climate data are 1951-1980 averages from PRISM (PRISM Climate Group, 2017).

| Region       | $R^2$ | P-value      | Predictor                 | Coefficient | t-statistic | P-value      |
|--------------|-------|--------------|---------------------------|-------------|-------------|--------------|
| Entire range | 0.42  | <b>0.001</b> | Geographic distance       | 0.0002      | 15.73       | <b>0.001</b> |
|              |       |              | Temperature differences   | 0.0028      | 1.44        | 0.486        |
|              |       |              | Precipitation differences | 0.0006      | 2.46        | 0.209        |
| North        | 0.36  | <b>0.008</b> | Geographic distance       | 0.0006      | 4.65        | <b>0.006</b> |
|              |       |              | Temperature differences   | 0.0061      | 1.63        | 0.377        |
|              |       |              | Precipitation differences | 0.0004      | 0.57        | 0.692        |
| Center       | 0.44  | 0.06         | Geographic distance       | 0.0006      | 4.87        | <b>0.001</b> |
|              |       |              | Temperature differences   | 0.0111      | 0.56        | 0.463        |
|              |       |              | Precipitation differences | -0.0005     | -0.45       | 0.737        |

Table 2 Covariance contributions of each layer in conStruct models with the number of layers (K) set to 1, 2, 3, 4, or 5.

| Number of layers    | 1     | 2      | 3      | 4      | 5      |
|---------------------|-------|--------|--------|--------|--------|
| Layer contributions | 1.000 | 0.9004 | 0.8062 | 0.8014 | 0.8925 |
|                     | -     | 0.0996 | 0.1043 | 0.1541 | 0.0795 |
|                     | -     | -      | 0.0895 | 0.0438 | 0.0204 |
|                     | -     | -      | -      | 0.0007 | 0.0055 |
|                     | -     | -      | -      | -      | 0.0021 |

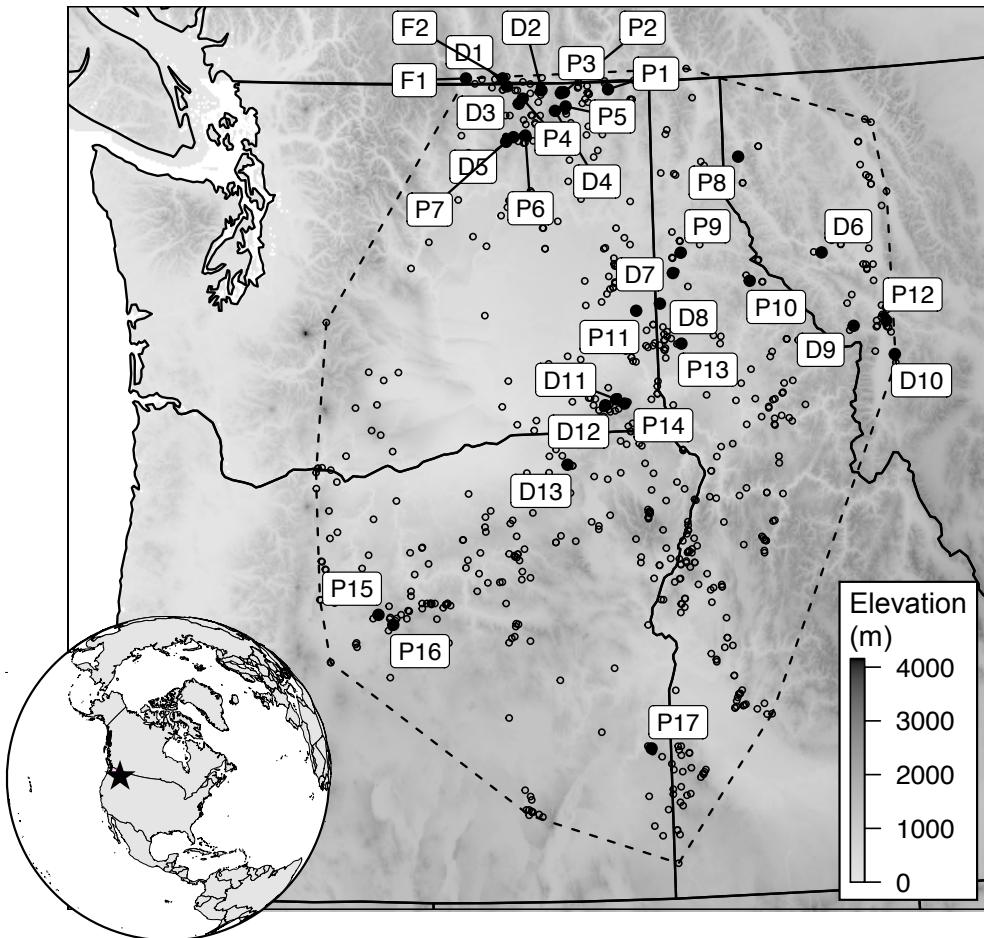


Figure 1 The geographic range of *Clarkia pulchella* across the interior of the Pacific Northwest. Small open points mark the locations of all herbarium records of *C. pulchella* from the Consortium of Pacific Northwest Herbaria that could be accurately assigned coordinates. The dashed line marks the maximum convex polygon drawn around these points. Larger filled points are populations that were sampled for this project. Labels correspond to population IDs in Table S1. Background shading shows elevation. The Columbia Basin is the unsampled area west of population D11.

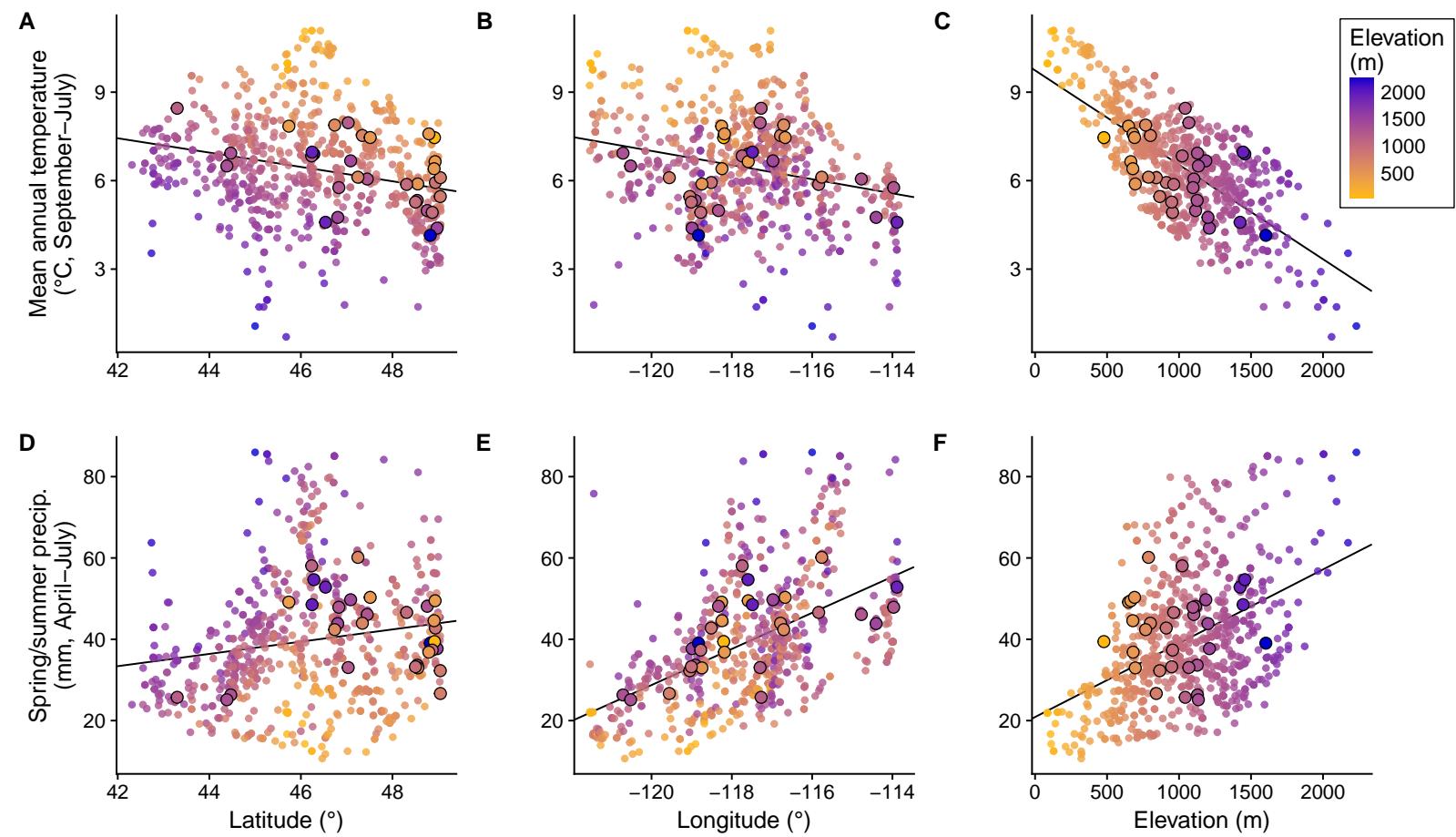


Figure 2 Relationships of climate and geography across the range of *Clarkia pulchella*. Small points represent all herbarium localities of *C. pulchella*, larger outlined points represent populations included in this study. Points are colored according to elevation. Temperature is influenced by (A) latitude, (B) longitude, and (C) elevation. Precipitation is also influenced by (D) latitude, (E) longitude, and (F) elevation. However, the interaction of these drivers results in climate that is heterogeneous across space. Climate data are 1951-1980 averages from PRISM (PRISM Climate Group, 2017). Trend lines are slopes from linear regression.

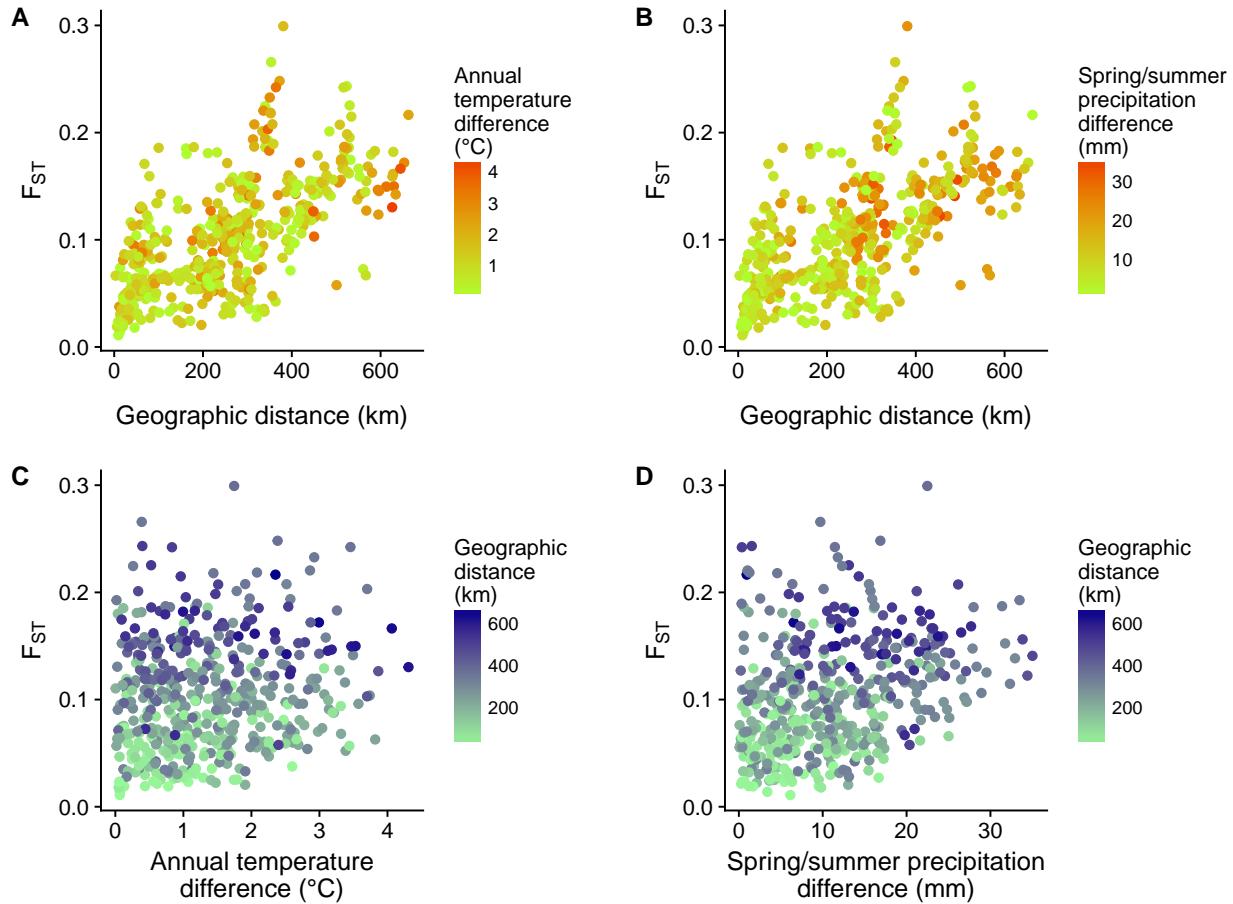


Figure 3 Pairwise genetic differentiation ( $F_{ST}$ ) of populations of *Clarkia pulchella* increases with geographic distance (x-axis in **A** and **B**), but shows no discernible relationship to temperature differences (color in **A**) or precipitation differences (color in **B**). An alternative visualization is presented in **(C)** and **(D)**, in which climate differences are plotted on the x-axis and geographic distance is indicated with color. Climate data are 1951–1980 averages from PRISM (PRISM Climate Group, 2017).

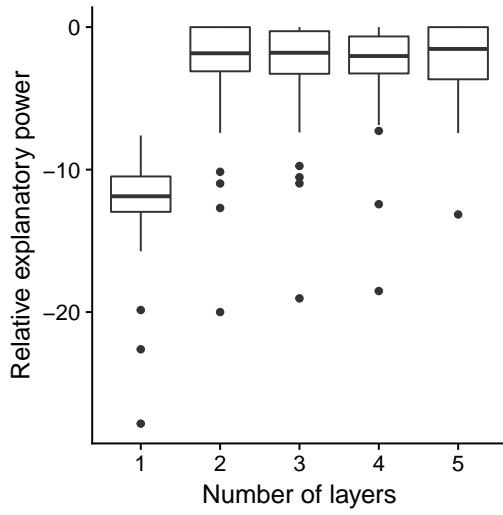


Figure 4 Results of 100 replicate cross-validation runs of conStruct with the number of layers set to 1, 2, 3, 4, or 5. In each replicate, the model is built using 90% of loci, and the log-likelihood of the remaining loci is calculated. Predictive accuracy is then calculated as the difference in log-likelihood between each model and the best model (i.e. the best number of layers) in each replicate. These results indicate that models constructed with two layers are best, because they provide as much explanatory power as other models without further complexity.

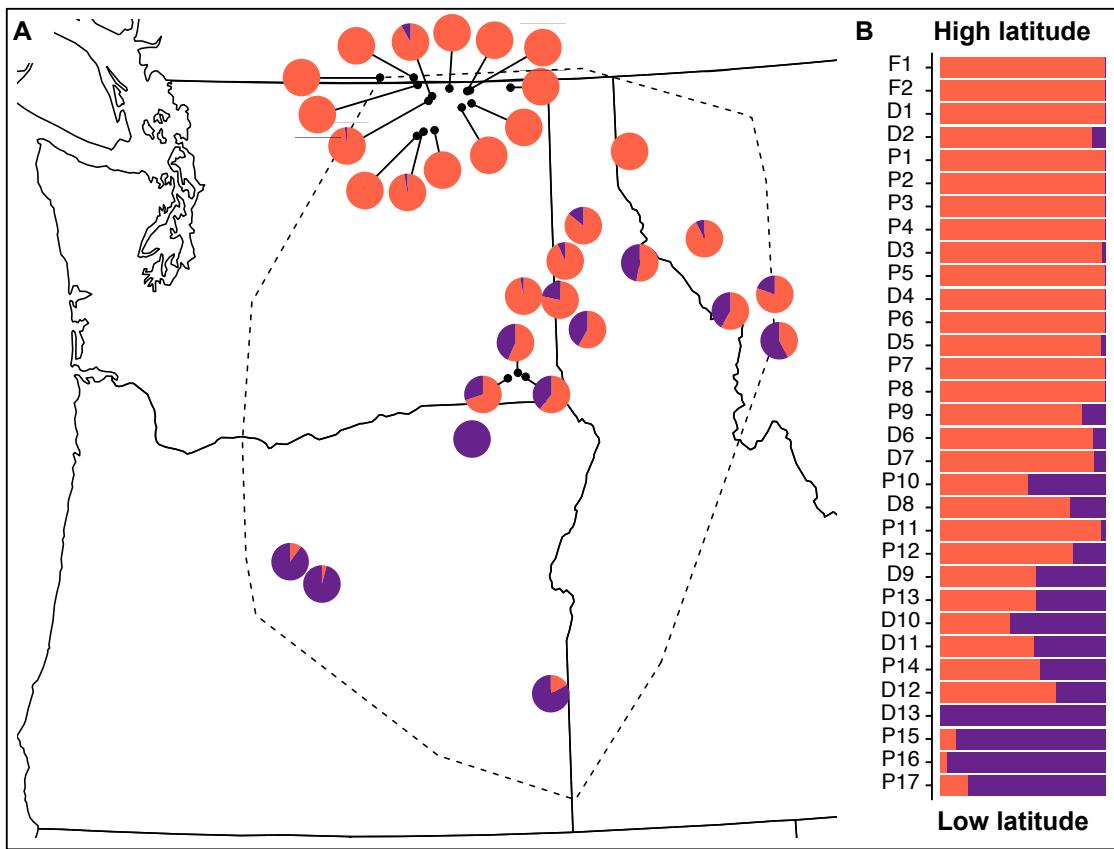


Figure 5 Admixture proportions of each of 32 populations of *Clarkia pulchella* estimated from by conStruct with  $K = 2$ . **A** Admixture proportions are shown in geographic space and **(B)** arranged by latitude . Population ID codes are consistent with Table S1 and Figure 1.

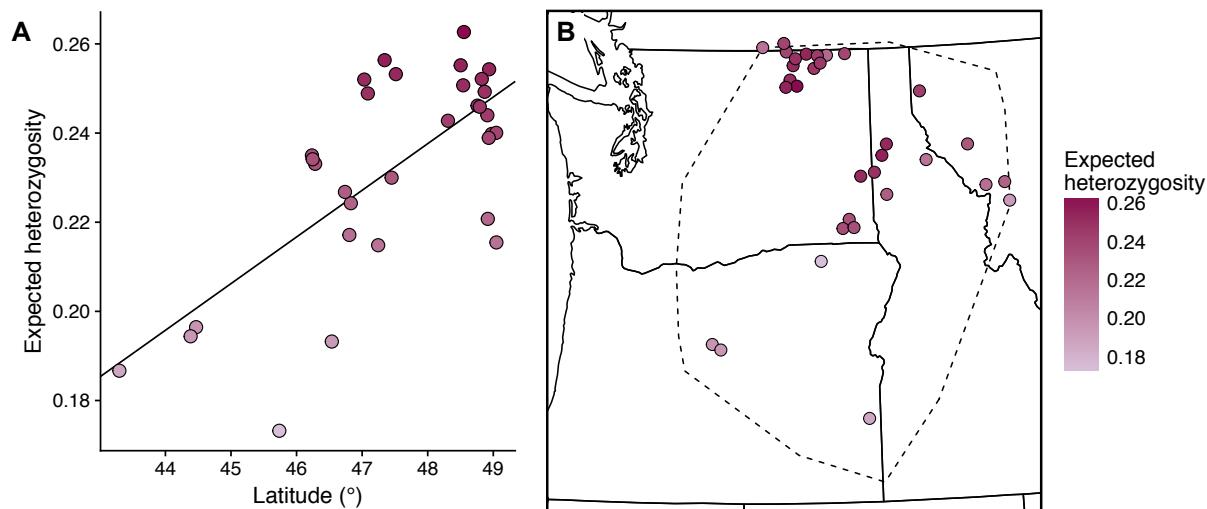


Figure 6 **(A)** Expected heterozygosity increases with latitude across the range of *Clarkia pulchella*. **(B)** Expected heterozygosity appears to be higher in central and northern parts of the range, but lower in the south and east.

Table S1 Geographic locations and elevations of populations of *Clarkia pulchella* included in these analyses. Population IDs are consistent with Figure 1. The populations included in analyses of geographic subsets are indicated.

| Population ID | Geographic subset | Latitude | Longitude | Elevation (m) |
|---------------|-------------------|----------|-----------|---------------|
| F1            | North             | 49.05    | -119.56   | 842           |
| F2            | North             | 49.04    | -119.05   | 866           |
| D1            | North             | 48.98    | -118.99   | 1211          |
| D2            | North             | 48.94    | -118.51   | 911           |
| P1            | North             | 48.93    | -117.59   | 665           |
| P2            | North             | 48.92    | -118.20   | 478           |
| P3            | North             | 48.91    | -118.25   | 679           |
| P4            | North             | 48.87    | -118.77   | 955           |
| D3            | North             | 48.83    | -118.83   | 1603          |
| P5            | North             | 48.79    | -118.18   | 681           |
| D4            | North             | 48.76    | -118.33   | 1115          |
| P6            | North             | 48.55    | -118.74   | 696           |
| D5            | North             | 48.54    | -118.91   | 1126          |
| P7            | North             | 48.50    | -119.01   | 949           |
| P8            | -                 | 48.31    | -115.84   | 963           |
| P9            | Center            | 47.51    | -116.67   | 691           |
| D6            | -                 | 47.45    | -114.77   | 1103          |
| D7            | Center            | 47.34    | -116.79   | 801           |
| P10           | -                 | 47.24    | -115.76   | 788           |
| D8            | Center            | 47.09    | -116.98   | 1186          |
| P11           | Center            | 47.03    | -117.30   | 1068          |
| P12           | -                 | 46.83    | -113.97   | 1097          |
| D9            | -                 | 46.80    | -114.41   | 1201          |
| P13           | Center            | 46.74    | -116.71   | 768           |
| D10           | -                 | 46.54    | -113.89   | 1424          |
| D11           | Center            | 46.28    | -117.60   | 1457          |
| P14           | Center            | 46.24    | -117.49   | 1445          |
| D12           | Center            | 46.24    | -117.74   | 1022          |
| D13           | Center            | 45.74    | -118.25   | 649           |
| P15           | -                 | 44.47    | -120.71   | 1128          |
| P16           | -                 | 44.38    | -120.52   | 1134          |
| P17           | -                 | 43.30    | -117.27   | 1043          |

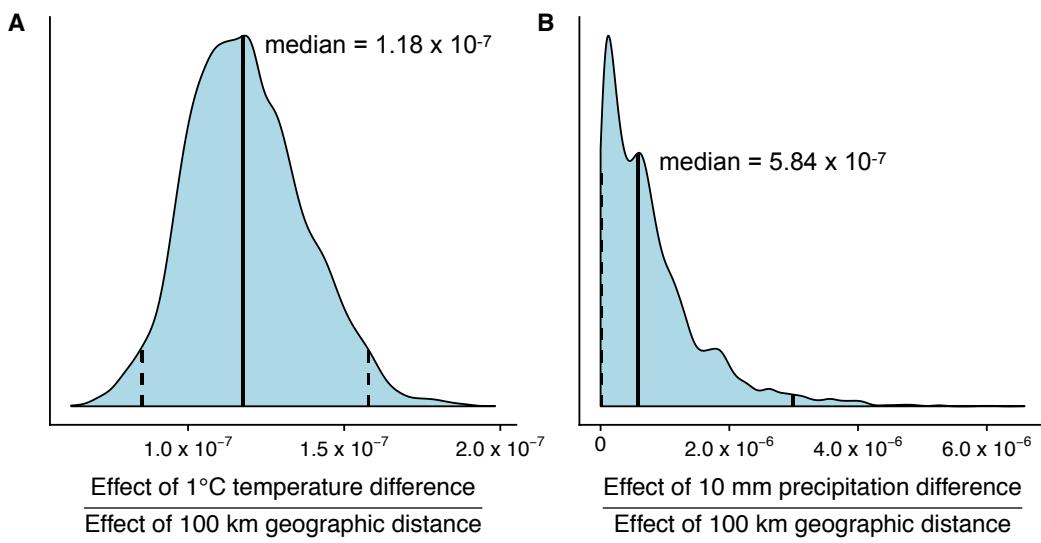


Figure S1 Marginal posterior distributions, median values (solid lines) and 95% credible intervals (dashed lines) of the ratio of the effect sizes of (A) temperature vs. geographic distance and (B) spring/summer precipitation vs. geographic distance on genetic differentiation of populations of *Clarkia pulchella* after a burn-in of 20%.

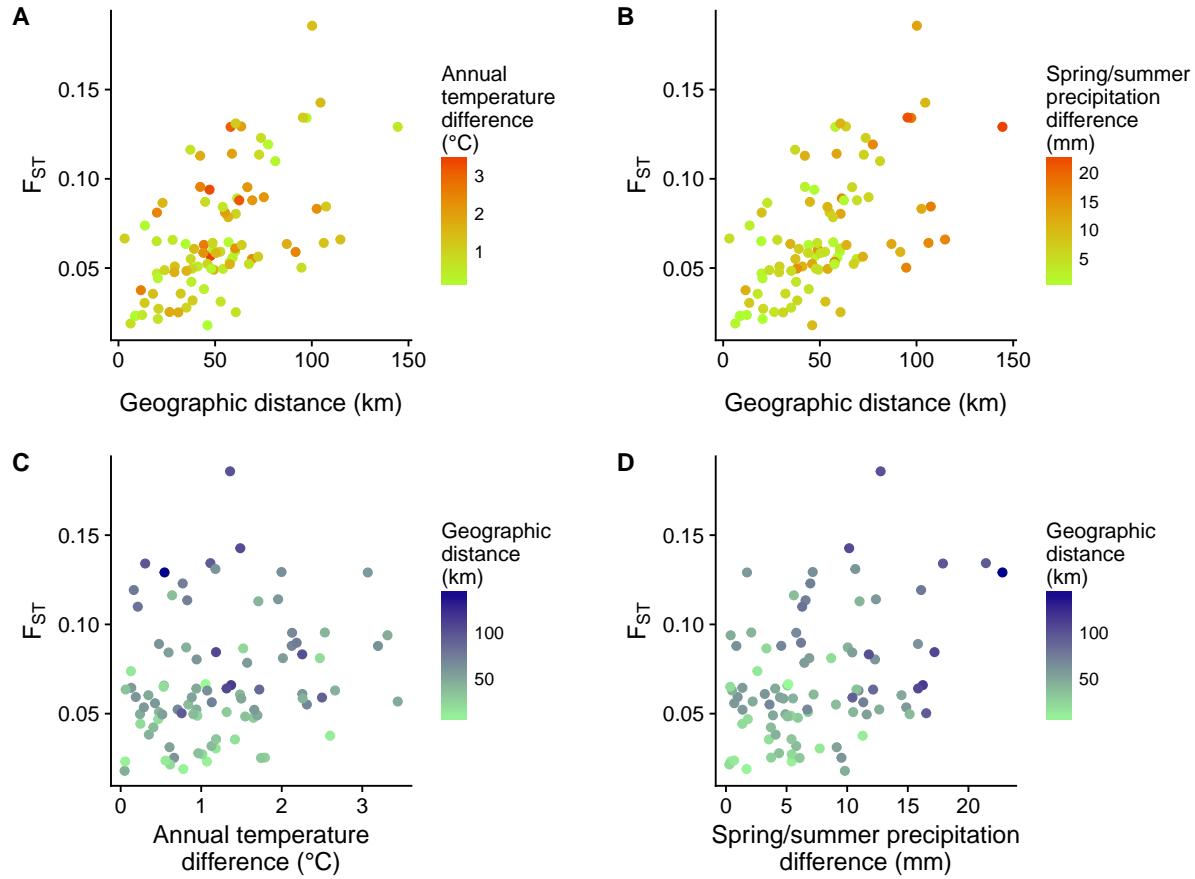


Figure S2 Relationship between pairwise geographic distance (x-axis in **A** and **B**), temperature differences (color in **A**) or precipitation differences (color in **B**), and genetic differentiation ( $F_{ST}$ ) among populations in the northern part of the geographic range of *Clarkia pulchella*. An alternative visualization is presented in **(C)** and **(D)**, in which climate differences are plotted on the x-axis and geographic distance is indicated with color. Climate data are 1951–1980 averages from PRISM (PRISM Climate Group, 2017).

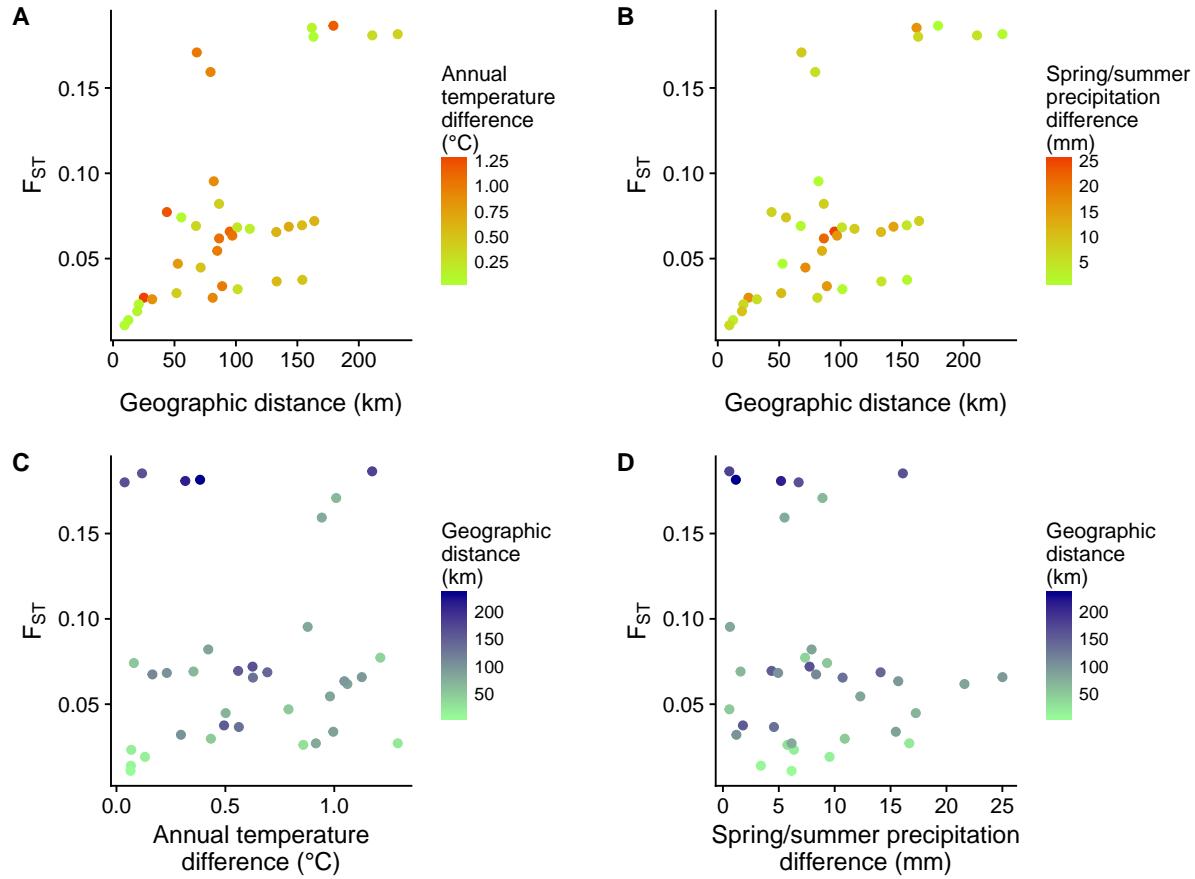


Figure S3 Relationship between pairwise geographic distance (x-axis in **A** and **B**), temperature differences (color in **A**) or precipitation differences (color in **B**), and genetic differentiation ( $F_{ST}$ ) among populations in the central part of the geographic range of *Clarkia pulchella*. An alternative visualization is presented in **(C)** and **(D)**, in which climate differences are plotted on the x-axis and geographic distance is indicated with color. Climate data are 1951–1980 averages from PRISM (PRISM Climate Group, 2017).