# Multi-level patterns of genetic structure and isolation by distance in the widespread plant Mimulus guttatus

# Multi-level patterns of genetic structure and isolation by distance in the widespread plant *Mimulus guttatus*

Alex D. Twyford[1,2,3], Edgar L. Y. Wong[2,4], Jannice Friedman[3,5]

[1]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK.

[2]Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh, UK.

[3]Department of Biology, Syracuse University, 107 College Place, Syracuse, New York, USA.

[4]Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, UK.

[5]Department of Biology, Queen's University, Kingston, Ontario, Canada.

Email for correspondence: Alex.Twyford@ed.ac.uk

Running title: The spatial scale of genetic structure in *Mimulus*

Word count: 5880

## Abstract

An understanding of genetic structure is essential for answering many questions in population genetics. However, complex population dynamics and scale-dependent processes can make it difficult to detect if there are distinct genetic clusters present in natural populations. Inferring discrete population structure is particularly challenging in the presence of continuous genetic variation such as isolation by distance. Here, we use the plant species *Mimulus guttatus* as a case study for understanding genetic structure at three spatial scales. We use reduced-representation sequencing and marker-based genotyping to understand dispersal dynamics and to characterise genetic structure. Our results provide insight into the spatial scale of genetic structure in a widespread plant species, and demonstrate how dispersal affects spatial genetic variation at the local, regional, and range-wide scale. At a fine-spatial scale, we show dispersal is rampant with little evidence of spatial genetic structure within populations. At a regional-scale, we show continuous differentiation driven by isolation by distance over hundreds of kilometres, with broad geographic genetic clusters that span major barriers to dispersal. Across Western North America, we observe geographic genetic structure and the genetic signature of multiple postglacial recolonization events, with historical gene flow linking isolated populations. Our genetic analyses show *M. guttatus* is highly dispersive and maintains large metapopulations with high intrapopulation variation. This high diversity and dispersal confounds

the inference of genetic structure, with multi-level sampling and spatially-explicit analyses required to understand population history.

## Introduction

Describing the pattern of genetic structure is the foundation for many population genetic studies. The cornerstone for understanding population genetic structure is the expectation that individuals become more genetically distinct, or less genetically related, with increasing geographic distance. Indeed, a pattern of "isolation by distance" is often used as a null model of genetic differentiation (Malecot 1948; Slatkin, 1993; Wright, 1943). This pattern arises because at increased distances, genetic drift and natural selection occur faster than dispersal can homogenise population differentiation. However, several challenges occur when studying population structure across species-wide distributions. In particular, the degree to which patterns represent individual dispersal and genetic drift versus large-scale population movements like recolonization from glacial refugia can generate patterns that are hard to interpret (Slatkin et al. 1987).

The development of new sequencing approaches make the study of natural populations more accessible (Ekblom and Galindo, 2011), and various methods allow genetic clusters to be detected and visualised across geographic space (Falush et al. 2003; Hubisz et al. 2009; Pritchard et al. 2000). However, the ease with which genetic structure can be detected is somewhat at odds with the known complexity of genetic variation in natural populations. Detecting clearly defined geographic genetic units is expected to

63 be challenging against the background of continuous genetic variation often
64 present in nature. As such, group assignment algorithms may detect artificial
65 genetic clusters in populations characterised by isolation by distance,
66 especially if the sampling of populations is aggregated (Pritchard et al. 2000).
67 Nevertheless, a recent meta-analysis has shown that isolation by distance is
68 present in 60% of population genetic datasets, with more than half these
69 studies continuing to use the programme STRUCTURE without accounting
70 for geographic distances between sampling locations (Perez et al. 2018). The
71 long-running issue of identifying 'clusters versus clines' motivated the
72 development of new methods that account for spatial information of samples
73 when estimating genetic structure (Bradburd et al. 2016; Petkova et al. 2016;
74 Bradburd et al. 2018; House and Hahn 2018), and these methods may be more
75 accurate in detecting distinct genetic clusters in the presence of continuous
76 patterns of genetic differentiation. However, only a few studies to date have
77 applied these methods (e.g. Murray et al. 2019; Whelan et al. 2019).
78
79 A further issue with studying genetic structure is going from pattern to
80 process, when a range of scale-dependent and temporally variable processes
81 together shape the spatial arrangement of genetic variation (Schregel et al.
82 2018). At a local scale, the amount of dispersal and the strength and pattern
83 of microgeographic selection, are expected to jointly affect spatial dynamics
84 and patterns of relatedness (Vekemans & Hardy 2004). At the population

level, barriers to dispersal and regional selection will determine metapopulation structure and consequently evolutionary dynamics (Husband & Spencer 1996). At the landscape-level, long-distance dispersal is expected to be uncommon, but even rare long-distance dispersal can introduce important allelic variation and affect broad-scale patterns of relatedness (Nathan 2006). Thus, studies that focus on a single spatial scale may bias their inferences toward a particular biological process. Therefore, to obtain a full understanding of the evolutionary and ecological processes that shape genetic variation and structure, studies should integrate over multiple scales (Schregel et al. 2018). However, such studies are rare.

Here, we investigate the spatial scale of genetic structure and address how dispersal interacts with other scale-dependent processes to determine genetic structure in the plant species *Mimulus guttatus* (syn. *Erythranthe guttata*; see Lowry et al. 2019 for nomenclature). While *M. guttatus* is a widely used study system for investigating diverse biological processes, from the evolution of flower colour pigmentation and patterns (e.g. Yuan et al. 2016; Twyford et al. 2018), to adaptation to harsh environments (e.g. Lowry et al. 2009; Hendrick et al. 2016), there are still major gaps in our knowledge of the structure of genetic variation in natural populations. Previous studies have shown populations of *M. guttatus* are strongly differentiated for adaptive traits and morphological characters (Friedman et al. 2015; Nesom 2014), and

107 population genetic divergence in this species is high ($F_{ST}$ ~0.5; Lowry et al.
108 2008; Puzey et al. 2017), therefore we may expect highly structured
109 populations, with localised genetic clusters and limited gene flow. In contrast,
110 owing to its high dispersal potential through seeds and vegetative fragments
111 (Lindsay 1964; Truscott et al. 2006; Vickery et al. 1986; Waser et al. 1982),
112 *M. guttatus* has spread to the farthest reaches of the Aleutian Islands in Alaska
113 and rapidly invaded large areas of north-western Europe and New Zealand
114 over the past c. 200 years (Truscott et al. 2006). Based on these observations,
115 one would predict broad-scale genetic structure but limited structure at a fine
116 spatial scale.
117
118 We address these different expectations by investigating genetic variation and
119 population structure at a range of spatial scales (Figure 1). First, we genotype
120 individuals at a fine-spatial scale to look for spatial genetic structure (SGS)
121 and limits to localised dispersal within two *M. guttatus* populations. Second,
122 we use genome-wide sequencing of populations across a 700 km transect
123 from the Sierra Nevada to infer the nature of genetic structure, using
124 conStruct (Bradburd et al. 2018), a recently developed method to infer
125 discrete genetic clusters from continuous population samples. Third, we
126 analyse sequence variation from range-wide populations to infer the extent of
127 divergence and the distribution of genetic variation, and to understand the
128 phylogeographic history of the species in North America. Finally, we

integrate across distance classes to understand dispersal dynamics over different scales. Our results provide critical insight into the genetic structure and phylogeographic history of a widespread and ecologically diverse plant species, while also allowing us to evaluate the benefits of studying dispersal at multiple spatial scales simultaneously.

## Materials and Methods

### Study species

The plant species *Mimulus guttatus* is an emerging model system in evolutionary and ecological research because of its rich adaptive variation, the presence of closely related interfile taxa, and its amenability to genetic analysis (Twyford et al. 2015; Wu et al. 2007). The species is a self-compatible hermaphrodite with small-flowered populations that are selfers or mixed-maters and large-flowered populations that are predominantly outcrossers. Pollination is by bees, and the small seeds are likely dispersed by wind and water. *M. guttatus* has two ecotypes, an annual ecotype found in seasonally dry conditions, and a perennial ecotype found in permanently wet sites (Lowry et al. 2008). These ecotypes show substantial morphological differentiation (Friedman et al. 2015), and are maintained by multiple regions of divergence, including a large chromosomal inversion that protects multiple loci involved in adaptive divergence (Lowry and Willis, 2010; Twyford and Friedman, 2015). Perennial populations reproduce vegetatively by producing

151 horizontal spreading stems (stolons), which may facilitate local clonal spread
152 and may also break-off and disperse along watercourses.

153

154 The extensive distribution range of *M. guttatus*, with a native range extending
155 over 5000 km from northern Mexico to Alaska, makes it a useful study system
156 for investigating geographic genetic structure and responses to biogeographic
157 barriers. *M. guttatus* is widespread and abundant in areas with a rich biota and
158 complex biogeography, encircling the Central Valley, spanning the
159 Cascade/Sierran transition, bridging the Cascades/Coast ranges and the
160 Rocky Mountains, and found in formerly glaciated regions of western Canada
161 and Alaska. Previous studies of *M. guttatus* populations have identified
162 geographic genetic structure corresponding to coastal and inland populations
163 (Lowry et al. 2008) or northern, coastal, and southern populations (Twyford
164 & Friedman 2015). However, the confounding issue of isolation by distance
165 observed in some population studies of *M. guttatus* (e.g. Kooyers et al. 2015),
166 and the lack of support for some nodes in phylogeographic analyses (Twyford
167 and Friedman, 2015), has precluded detailed interpretation of geographic
168 genetic structure and the phylogeographic history of the species.

169

170 **Fine-scale spatial genetic structure**

171 We used two populations from California to estimate spatial genetic structure.
172 We sampled one population of the perennial ecotype (population ELD), and

one of the annual ecotype (FOR; see Table S1 for population details). Our classification of population life history was based on morphological traits such as number of stolons and flower size, observed in the field and in common garden experiments (Twyford & Friedman, Unpublished data). The two study populations were chosen for their large census population sizes of many thousands of individuals, and the continuous distribution of individuals with no obvious barriers to dispersal. For each population, we sampled at least twenty plants at approximately 30 cm intervals along a transect, with four additional transects at different spacings (3 - 500 m). Our sampling scheme represents a shallow survey of individuals, sampling less than 10% of plants in the populations, with the aim to capture the range of pairwise distance classes represented by samples within each site. Maximum inter-plant sampling distances were 680 m for population ELD and 410 m for population FOR. We calculated interplant distances from individual GPS coordinates. We collected plant tissue in silica gel, for DNA extraction with the Qiagen Plant DNeasy kit (Qiagen, Germantown, MD). We used a total of ten PCR-based markers for genetic analysis, four intron-based length polymorphism markers and six microsatellites (marker details reported in Lowry et al. 2008). We performed multiplexed PCR reactions with M13-tailed primers, prior to genotyping on the ABI 3730 DNA Analyzer at Edinburgh Genomics. We scored the size of the amplified fragments automatically, with manual edits, using geneMapper (Applied Biosystems). We checked genotype data for null

195 alleles and other errors with MICRO-CHECKER (Van Oosterhout et al.
196 2004). We excluded locus MgSTS278 in population ELD due to uneven
197 amplification success, prior to statistical analyses. The final dataset included
198 91 individuals for population ELD and 79 for FOR.
199
200 We inferred individuals likely to be the product of clonal reproduction using
201 a permutation and re-sampling approach that accounts for scoring error and
202 somatic mutations, as implemented in GENCLONE (Arnaud-Haond &
203 Belkhir 2007). We related clonality to the inter-plant sampling distance to
204 understand the extent of clonal spread, then selected a single individual at
205 random from each clone for downstream analyses of diversity and
206 relatedness. We calculated the extent of SGS for each population using spatial
207 autocorrelation analysis described in Vekemans and Hardy (2004), using the
208 pairwise kinship coefficients (F*ij*) of Loiselle et al. (1995). We performed
209 analyses with SPAGeDi (Vekemans & Hardy 2004) using the following
210 distance classes: 0 – 2 m, 2 – 4 m, 4 – 6 m, 6 – 8 m, 8 – 10 m, 10 – 20 m, 20
211 – 50 m, 50 – 100 m, 100 – 200 m, 200 – 400 m and 400 – 700 m. We
212 calculated mean F*ij* per distance class, 95% confidence intervals by
213 permutation, standard errors by jack-knifing, and plotted autocorrelograms
214 for each analysis. We calculated overall spatial genetic structure per
215 population with the *Sp* statistic. As we found little evidence of genetic
216 substructure within populations (see results), we then calculated pooled

diversity statistics across transects within a site. We used FSTAT v.2.9.3 (Goudet 2001) to calculate the inbreeding coefficient ($F_{IS}$) and allelic richness ($A_R$) per population.

**Population-level differentiation**

We calculated the extent of population-level genetic diversity and differentiation for a transect of nine populations of *M. guttatus* spaced at approximately 95km intervals (range 52-143 km) through the Sierra Nevada (Table S1). We collected leaf tissue from between eighteen and twenty well-spaced (>1 m) individuals per population into silica for DNA extraction. We used the genotyping by sequencing (GBS) method to generate genome-wide polymorphism data (Elshire et al. 2011). We created sequencing libraries by digesting individual samples with the frequent cutting enzyme ApeKI, before ligating barcoded adapters, performing PCR, and pooling in 96-plex reactions. We sequenced multiplexed libraries with 100 bp single-end sequencing with the Illumina HiSeq 2500 at Rochester Medical Center. We used TASSEL-GBS v2 (Glaubitz et al. 2014) to de-multiplex samples, remove barcodes, perform quality filtering, and call SNPs. We aligned the GBS tags to the *M. guttatus* genome version 2.0_256 (phytozome.net) using the default settings of BWA (Li & Durbin 2009). We called sites with a minimum quality score of 20, and with no minimum allele frequency to recover all variant and invariant sites. Ten of 193 sequence libraries failed,

239 yielding less than 1% of the mean number of sequencing reads across
240 samples, while other samples yielded between 766,539 – 6,704,555 reads.
241 The average sequencing coverage per site was 38-fold, for 5,611,458 sites.
242 Downstream population genetic analyses used a subset of data filtered to
243 include individuals with less than 50% missing data, sites scored in over 75%
244 individuals, a minor allele frequency of 0.05, and with SNPs in tight linkage
245 removed by filtering variants within 20 bp (Brandvain et al. 2014), to give a
246 final dataset of 22,697 SNPs.
247
248 We inferred discrete population structure using conStruct (Bradburd et al.
249 2018), which models admixture across a specified number of discrete layers
250 as defined by the *K*-value. Non-spatial conStruct analyses do not use location
251 information, while spatial conStruct analyses assume allele frequencies have
252 a positive covariance based on geographic locations to account for isolation
253 by distance. To determine an appropriate level of parameterization for the
254 models, we used cross-validation with a training set (Bradburd et al. 2018),
255 and compared predictive accuracies between spatial and non-spatial models,
256 and between successive *K*-values, to determine which model has the best
257 goodness-of-fit without overfitting. We analysed *K*-values of 1-9. To test
258 whether spatial models were better fitting than non-spatial models we used
259 paired *t*-tests comparing cross-validation scores across values of *K*. The best
260 fitting models were repeated with 100,000 MCMC iterations with the first

50% removed as burn-in to produce the final analyses. Admixture plots were visualised per population using the default options in conStruct.

We compared our conStruct results with patterns of genetic structure inferred from additional non-spatial analyses. fastSTRUCTURE analyses used the simple prior and values of *K* between 1 and 9, with the optimal *K* considered as the run that maximizes the log-marginal likelihood of the data. We then re-ran fastSTRUCTURE with the logistic prior, to help infer fine-scale admixture. Admixture plots were visualised per individual using the default options in fastSTRUCTURE. We performed PCA analysis in Tassel (Bradbury et al. 2007) and calculated pairwise $F_{ST}$ using the R package diveRsity (Keenan et al. 2013) and nucleotide diversity ($\pi$) per site (including invariant sites) using VCFTools (Danecek et al. 2011).

**Range-wide dispersal and broad-scale genetic structure**

We reanalysed GBS data from 174 individuals from 70 populations from across the native range of *M. guttatus* which were used to compare SNP differences within and outside a chromosomal inversion by Twyford & Friedman (2015). This data includes annual and perennial populations sampled from Alaska, Arizona, California, Idaho, Nevada, Oregon and Washington (America), as well as British Columbia (Canada) and Sonora (Mexico) (Figure S1). We re-called SNPs from the raw reads using the Tassel

5 GBS v2 pipeline, with the minor allele frequency set to 0 to call invariant sites to improve branch length estimates in phylogenetic analyses. Variant calls were made using a minimum sequencing quality score of 20. Our variant calling produced 72,941 SNPs and invariant sites that were used in phylogenetic analyses, of which 6,523 sites were variable. Two further filtered datasets were generated for population genomic analyses. For analyses of genetic structure, we filtered invariant, low frequency sites and SNPs in tight linkage (as above), and removed samples with more that 25% missing data, producing a dataset of 3,414 SNPs. For TreeMix analysis, we filtered populations with fewer than three sampled individuals, leaving 30 populations, and then filtered invariant, low frequency sites and SNPs in tight linkage as above, to give a final dataset of 3,066 filtered SNPs.

We used conStruct, as described above (but with *K*-values between 1 and 10), to characterize genetic structure using spatial and non-spatial models. We then used polymorphism-aware phylogenetic models (PoMo) implemented in IQ-TREE (Nguyen et al. 2015) to investigate population-level relationships. PoMo uses site frequency data to account for incomplete lineage sorting thus providing a more accurate estimate of the species tree when there is gene discordance (De Maio et al. 2015). We calculated allele frequencies per population using the counts file library (cflib) python scripts supplied with IQ-TREE. We tested the best-fitting model (-m TEST) and subsequently

performed analyses with TVM+F+G4+P. We adjusted the virtual population size setting (N) to equal the number of chromosome sets per population (i.e. +N5) based on the mean of 2.5 diploid individuals sampled per site. Tree searches used settings recommended for short-sequence block data (-pers 0.2, -nstop 500). We used 1000 ultrafast bootstrap estimates to test the support for the topology (Minh et al. 2013).

We used TreeMix to further investigate population relationships and to model historical migration events. TreeMix constructs a maximum likelihood phylogeny from genome-wide polymorphism data, and incorporates directional migration edges between populations where historical admixture is likely (Pickrell & Pritchard 2012). We assessed the fit of models with between 0 and 10 migration events by calculating the percentage of variation explained by the maximum likelihood trees using the treemixVarianceExplained scripts as part of the RADpipe package (doi: 10.5281/zenodo.17809). We also investigated patterns of range-wide genetic diversity by calculating $\pi$ per site for each population with two or more sampled individuals, using VCFTools. We used general linear models in R to test whether variation in $\pi$ is explained by life history (annual vs. perennial) and geographic region (coastal, northern, southern).

**Integrated analyses across spatial scales**

We evaluated the pattern of genetic structure across spatial scales. First, we evaluated the strength of isolation by distance by regressing pairwise population genetic structure ($F_{ST}/(1-F_{ST})$) against pairwise linearised geographic distance (log transformed). We did this separately for the Sierra dataset and the range-wide data. We tested for a correlation between the matrix of geographic distances and the matrix of genetic distances using a Mantel test with 99 permutations in the R package Ade4 (Dray & Dafour 2007). Next, we used the geostatistical method of using semivariance to fit variograms to our genetic divergence and geographic distance data to understand broad-scale patterns of genetic relatedness. We performed analyses separately for the Sierra and range-wide data. We fitted variograms using the R package Phylin (Pedro et al. 2015) with the 'gen.variogram' function, and models with the 'gv.model' function. We permuted the nugget and sill to identify the best-fit model measured by $R^2$ fit to the data.

**Results**

Our genotyping of spatially mapped individuals in a population of the annual ecotype (FOR) revealed high mean allelic richness ($A_R$ = 10.4), with low overall SGS (*Sp* = -5.90 x $10^{-5}$). Over all distance classes, values of F*ij* consistently fell within the permuted upper and lower confidence intervals, reflecting no spatial structure (Figure 2A). A population of the perennial ecotype (ELD) also showed high genetic diversity as measured by mean

allelic richness ($A_R$ = 8.1), though there was evidence for local clonal spread, with two to six samples present in nine clonal genotypes, with a maximal clonal spread of 4.8 m. There was no SGS in most distance classes, except a high and significant F*ij* value in the 0 - 2 m distance class (Figure 2B). Both populations also had evidence of non-random mating, with a high $F_{IS}$ value for the annual population ($F_{IS}$ = 0.388) and a moderate $F_{IS}$ value in the perennial population ($F_{IS}$ = 0.218). Overall, the general absence of SGS suggests no limits to dispersal over a spatial scale of hundreds of meters in large continuous *M. guttatus* populations, though clonal spread and self-fertilisation influence fine-scale population dynamics.

Genome-wide SNP analysis of nine *M. guttatus* populations spaced at ~100 km intervals through the Sierra Nevada showed high genetic diversity with a mean per site $\pi$ of 1.6 %, and high population structure with a mean pairwise $F_{ST}$ of 0.327. Analyses of genetic clustering using PCA and fastSTRUCTURE revealed geographic genetic clusters corresponding to northern and southern Sierran populations (Figure 3). Similarly, the non-spatial model in conStruct showed a north-south genetic division at $K$ = 2 (Figure 4A), with genetic clusters corresponding to geographic groupings at $K$ = 3 or 4 (Figure 5A-C; with $K$ = 4 the value at which the likelihood plateaus, Figure S2). However, model-based clustering incorporating spatial information proved a significantly better fit than non-spatial models, particularly for *K*-values

between one and three (Figure S2). The spatial conStruct models did not show a clear north-south genetic discontinuity at $K = 2$ (Figure 4C), and at higher $K$-values (Figure 5D-F), and instead genetic differentiation of the Sierra Nevada populations largely reflects continuous variation in allele frequencies rather than discrete genetic clusters.

Analyses of broad-scale population samples across the native range, using non-spatial conStruct models, revealed clinal genetic variation at $K = 2$ that correlates with latitude ($R^2 = 0.545$; Figure 4B, 5G). This result confirms a pattern of south-north genetic divergence previously identified with STRUCTURE (Twyford and Friedman 2015). In contrast, spatial conStruct models showed a substantially better fit to the data across $K$-values ($P < 0.0001$, Figure S2), revealed no such correlation with latitude at $K = 2$ ($R^2 = 0.0001$), and instead discriminated coastal from inland populations (Figure 5J). At $K = 3$, spatial models distinguished coastal, northern and southern genetic clusters but with major admixture (Figure 5K).

Complex patterns of genetic structure were also evident in polymorphism aware phylogenetic analyses, which resolved a well-supported tree topology with genetic clusters of southern, coastal and northern populations, while also revealing previously uncharacterised geographic substructure within clades (Figure 6). For example, well-supported at the base of the coastal clade are

two Californian populations from Monterey Bay, BCB and LOR, with other more northerly coastal populations in a derived position, supporting south to north range expansion along the Pacific coast. Evidence for historical dispersal in *M. guttatus* is provided by the TreeMix analysis, with models incorporating at least two migration events (*m*) showing much better model fit than those without migration (Figure S3). At $m = 10$, dispersal is observed across the admixture graph, including multiple dispersal events from populations in California (Figure 6C). These results indicate a history of repeated dispersal across the range of *M. guttatus,* facilitating recolonization after glaciation.

Genetic diversity as estimated by population-level $\pi$ values showed no significant difference between annual and perennial populations ($F_{1,55}$=0.22, $P = 0.6$), and instead the three previously identified geographic clusters explain a significant amount of variation in the data ($F_{2,55}$=5.75, $P < 0.01$). The greatest genetic diversity was found in the southern cluster, then northern, and the lowest in the coastal cluster (Figure 6A). While genetic diversity was uniformly low across populations along the 580 km of coastline in Oregon, and uniformly high across the north of the Sierra Nevada, genetic diversity was more heterogeneous in other areas, with notable patches of high diversity both in the north (e.g. HOC, Olympic National Forest, Washington)

414 and in the more sparsely sampled inland southern populations of Sonora
415 (ALI) and Arizona (CRZ).
416
417 For the Sierra dataset, the linear regression between ($F_{ST}$ /1-$F_{ST}$) and log
418 pairwise geographic distance was significant ($P < 0.001$) and geographic
419 distance explained 66% of genetic variation (Figure 7A). However, for the
420 range-wide dataset, although the linear regression is significant ($P < 0.001$),
421 the data showed a poor fit, with distance explaining only 6% of genetic
422 variation (Figure 7B). This matches predictions that isolation by distance
423 should break down with increasing geographic distance as dispersal processes
424 change and as different geographic genetic clusters mix. Similarly, Mantel
425 tests for the correlation between the geographic and genetic distance matrices
426 showed a significant correlation for the Sierra dataset ($r$: 0.69, $P < 0.01$), but
427 no significant correlation for the range-wide dataset ($r$: 0.13, $P = 0.13$). To
428 provide a separate estimate of the geographic scale of genetic differentiation
429 independent of mutation rate we related genetic divergence to geographic
430 distance in variogram models. Our results showed that the range, defined as
431 the scale of spatial autocorrelation after which little change in the semi-
432 variance is encountered with increasing distances, extended to 500 km for
433 both Sierra and range-wide GBS data (Figure 7C, D). These analyses showed
434 that the spatial independence of populations is only achieved at 500 km.
435

## Discussion

Identifying the spatial scale of genetic structure is key for understanding population dynamics and for inferring evolutionary and ecological processes, however most population genetic studies focus on a single spatial scale or ignore spatial information in their analyses. Our analyses of the widespread plant *M. guttatus* revealed different patterns of genetic structure over a range of spatial scales. Within populations, we observed a lack of spatial genetic structure, suggesting extensive local dispersal. Between populations, we identified continuous genetic variation and isolation by distance, which had a major impact on the inference of genetic clusters. After accounting for isolation by distance, we were able to distinguish broad geographic genetic clusters that spanned many well-characterised barriers to dispersal. Across the species' native range, we observed geographic genetic clusters corresponding to repeated colonisation from the south, with evidence for widespread historical dispersal. This pattern of recurrent colonisation suggests the species is an excellent coloniser that rapidly expands its range in response to new ecological opportunities and habitat availability. Our results showing high diversity and broad-scale genetic structure support the finding that *M. guttatus* has large metapopulations with high intrapopulation variation (Puzey et al. 2017). Local genetic variation and genetic structure is shaped by diverse factors including self-fertilisation and clonal spread of the perennial ecotype, in conjunction with diverse forms of selection known to operate in

458 this species (Peterson et al. 2016; Troth et al. 2018). We discuss our results
459 below in terms of the spatial scale of gene flow and the species' historical
460 demography, and make recommendations for how best to use genetic
461 information to infer genetic structure at different spatial scales.
462
463 **Spatial dynamics of dispersal and migration**
464 Our findings show that high dispersal potential has shaped genetic structure
465 of *M. guttatus* populations. The lack of spatial genetic structure over hundreds
466 of meters likely reflects extensive local dispersal. Similarly, the emergence
467 of broad geographic genetic clusters that extend over potential barriers to
468 dispersal, and variogram analyses that reveal the non-independence of
469 populations over hundreds of kilometres, suggests large metapopulations.
470 Although pollen movement by bees, and downstream dispersal of vegetative
471 fragments will contribute to dispersal in this species, seed-mediated dispersal
472 is likely to dominate. While >40% of *M. guttatus* seeds fall within 25 cm of
473 the maternal plant (Ritland & Ritland 1996; Sweigart et al. 1999; Vickery et
474 al. 1986), giving rise to some localized fine-scale genetic structure (Ritland
475 and Ritland 1996) as observed in the perennial population, many of the
476 lightweight seeds (0.002 mg) are likely dispersed much further. Occasional
477 long-distance seed dispersal by wind, animals or water (Martin 2004) may be
478 crucial for the widespread colonisation of *Mimulus* in its native range, and
479 also in its introduced range where it has become a dominant species of

480 disturbed watercourses over the last 200 years. Taken together, our analyses
481 and previous work show the important role of dispersal at all spatial scales.
482
483 Research on genetic population structure in natural populations is often
484 focussed on assessing genetic divergence (e.g. $F_{ST} \sim 1/(4N_em+1)$) and
485 understating demographic connectivity by predicting migration rate ($m$).
486 However, linking $F_{ST}$ and $m$ relies on a number of assumptions that may be
487 unrealistic in most natural populations. For example, the challenge for
488 understanding migration is illustrated by the contrast between high $F_{ST}$ values
489 in *M. guttatus* which imply low migration (this study, and others reviewed in
490 Puzey et al. 2017), and a migration rate sufficient to homogenize population
491 differences in models fit to whole genome data (Aeschbacher et al., 2017).
492 Crucial to estimating migration is understanding diversity and the effective
493 population size ($N_e$), with previous estimates of $N_e$ for *M. guttatus* in the
494 hundreds of thousands (between $4.805 \times 10^5$ and $6.730 \times 10^5$: Aeschbacher
495 et al. 2017; Brandvain et al. 2014), while the synonymous $\pi$ value of 3.3%
496 estimated by Puzey et al. (2017) makes *M. guttatus* one of the most
497 genetically diverse plant species studied to date. Our study supports the
498 finding of high genetic variation maintained in *M. guttatus* populations—we
499 found up to thirty alleles at polymorphic markers within a population,
500 numerous unique genotypes in a clonal perennial population, and a high value
501 for sequence diversity at 1.6%. High genetic diversity was present even in

502 populations with moderate to high selfing rates. Nonetheless, the demography
503 of *M. guttatus* populations are characterized by 'boom and bust' dynamics,
504 with rapid colonization and population expansion in response to ecological
505 opportunities, but with frequent local extinctions due to drought and habitat
506 change (Vickery 1999). Overall, it is possible that while seed dispersal allows
507 substantial mixing within populations and facilitates occasional long-distance
508 dispersal, migration between populations is not always sufficient to
509 homogenize population differences.
510
511 **Spatial scale of genetic structure**
512 Our work highlights the confounding influence of continuous genetic
513 variation on the inference of genetic clusters. At regional spatial scales, for
514 example across Sierra Nevada populations of *M. guttatus*, we found strong
515 isolation by distance, and spatial analyses accounting for continuous
516 population structure did not detect clear geographic genetic structure. This
517 indicates continuous genetic variation with geographic clines in allele
518 frequencies, rather than discrete population clusters due to barriers to
519 dispersal. Nonetheless, there was some evidence of subtle north-south
520 divergence in the Sierra Nevada, which is notably less distinct than in other
521 organisms. In spiders, for example, cryptic intraspecific breaks and species
522 divergence were found between Sierran populations (Hedin et al. 2013).
523 Interestingly, we found contrasting patterns between analyses of genetic

structure across the species range. Non-spatial analyses such as fastSTRUCTURE and PCA detected the genetic distinctiveness of sampling sites, which in *M. guttatus* correspond to south-north genetic structure, while spatial analyses such as conStruct revealed clusters corresponding to barriers to dispersal and demographic history, which are coastal and non-coastal populations in *M. guttatus*. This underscores that spatial and non-spatial analyses complement each other and reveal different aspects of population structure (Bradburd et al. 2018).

**Demographic and phylogeographic history**

Combining our analyses of genetic structure and phylogeography allow us to suggest a model for the historical colonisation of *M. guttatus* across the US. The joint evidence from the phylogeographic and genetic diversity analyses support southern populations as a reservoir of diversity and a major source for range expansion. Divergence of populations in the south of the species range is likely to have occurred in the Pleistocene around 265,000 years ago (Brandvain et al. 2014). This postdates the period of major geological uplift during the Pliocene (3–5 Ma), or pre-Pliocene activity, and instead supports glacial activity and consequent climatic changes in the Sierra Nevada structuring genetic diversity in *M. guttatus*. The location of inland refugia is hard to specify due to the uniformly high genetic diversity of these populations, however this seems most likely to be in the south of the Sierra

546 Nevada. Whether there was a separate coastal refugium is hard to say with
547 certainty. A coastal refugium is recognized for many North American plant
548 species (Brunsfeld et al. 2001), and the patchy occurrence of high genetic
549 diversity in coastal populations, and the topology of the phylogeny, are
550 broadly consistent with a separate coastal refugium at the southern extent of
551 the Northwest Forested Mountain biogeographic area, in the region of the
552 Wilson Grove Formation. However, evidence for shared genetic variants
553 between coastal populations and a population approximately 200 km inland,
554 East of the Central Valley (population MED), suggests coastal populations
555 may be independently derived from an inland source, rather than from a
556 separate coastal refugium. Similar patterns of trans-valley relatedness have
557 been seen in spiders (Hedin et al. 2013) and salamanders (Reilly et al. 2015).
558 Major rivers are orientated in a perpendicular axis to the Sierra Nevada
559 mountains (Rovito, 2010), and may have acted as a route for dispersal of the
560 perennial ecotype. Our results support the model of Western North American
561 phylogeography proposed by Brunsfeld et al. (2001), where vicariance,
562 dispersal and refugia shape genomic variation, and where dispersal has
563 occurred in waves as postglacial conditions became more hospitable.
564
565 Subsequent range expansion from glacial refugia has left a clear genetic
566 signature, with a latitudinal cline of genetic variation across the north of the
567 species range that parallels broad-scale north-south divergence seen in other

568 taxa such as wild sunflowers (McAssey et al. 2016). Northwards range
569 expansion and a subsequent increase in population size is likely to be recent,
570 within the last ~20,000 years (Brandvain et al. 2014), as a response to
571 increased habitat availability and more hospitable conditions post-glaciation.
572 The improved branch support in polymorphism-aware phylogenetic analyses
573 compared with conventional Bayesian analysis of concatenated sequences
574 (Twyford and Friedman, 2015), supports the divergence of northern and
575 coastal populations, and suggests a scenario of multiple independent
576 colonisation events from the south, each with different biogeographic
577 histories. Of particular interest is the coastal genetic cluster, which is mostly
578 restricted to a narrow band adjacent to the Pacific. The low genetic diversity
579 suggests these populations have been through a genetic bottleneck, while the
580 TreeMix analysis suggests these populations have subsequently been a major
581 source of admixture with inland populations. Overall, range-wide genetic
582 variation in *M. guttatus* has been shaped by recurrent colonisation from the
583 south of the species range, with dispersal avenues facilitating colonisation.
584
585 Our finding of multiple independent recolonization events has important
586 consequences for selecting samples for demographic analyses of *M. guttatus*.
587 Genomic studies using sparse population samples need to compare
588 individuals derived from a similar range expansion event, otherwise
589 demographic inferences will reflect ancestral variation rather than recent

population changes. While our data support (at least) two colonization events from the south of the species range, recolonization from a northern refugium, such as the Berengian refugia proposed for cold-tolerant taxa such as the serrated wintergreen *Orthilia secunda* (Beatty & Provan 2010), generally seems less likely for *M. guttatus* given the placement of northern populations as highly derived in the population phylogeny. However, there are patches of high genetic diversity in the north, and TreeMix shows these populations are both a sink and a source of migration. The question of cryptic northern refugia would be better resolved with detailed sampling from the north of the species range.

## Conclusion

Our genetic analyses reveal how dispersal affects spatial genetic variation from the local, to the regional, to the range-wide scale. At a local scale, high dispersal interacts with factors such as the spread of clonal genotypes and inbreeding, while at the broad spatial scale genetic structure is more likely to be determined by historical demography. Studying a single spatial scale would have overlooked critical aspects of metapopulation structure and limited our ability to infer dispersal dynamics, while not using spatial analyses would have overestimated the extent of geographic genetic structure where there is strong isolation by distance. We recommend other studies of population structure combine genetic data at multiple spatial scales, as well

612 as make use of spatial analyses of genetic structure to better understand
613 genetic variation in widespread species.
614
615 **Acknowledgements**
616 We thank Simon Aeschbacher for useful discussions of models of gene flow,
617 Gideon Bradburd for advice on conStruct, Hannes Becher for help with R
618 code, Jaanus Remm for assistance with interpreting variograms and Mabon
619 Elis for help with figures. This research was supported by a Natural
620 Environment Research Council Fellowship (NE/L011336/1) and a Heredity
621 Fieldwork Grant to ADT, and a National Science Foundation grant (DEB-
622 1354259) to JF. The fine-scale spatial genetics component formed part of the
623 MSc thesis of ELYW at the RBGE, which is funded from the Scottish
624 Government's Rural and Environment Science and Analytical Services
625 Division (RESAS).
626
627 **Conflicts of interest**
628 The authors declare that they have no conflict of interest.
629
630 **Data Accessibility Statement**
631 The raw sequence reads are available in the SRA, and marker genotypes and
632 aligned SNP data in Dryad (DOI on acceptance).
633

## Literature cited


Aeschbacher S, Selby JP, Willis JH, Coop G (2017) Population-genomic inference of the strength and timing of selection against gene flow. *Proceedings of the National Academy of Sciences* 114:7061-7066.

Arnaud-Haond S, Belkhir K (2007) GENCLONE: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Molecular Ecology Notes* 7:15-17.

Beatty GE, Provan J (2010) Refugial persistence and postglacial recolonization of North America by the cold-tolerant herbaceous plant *Orthilia secunda*. *Molecular Ecology* 19:5009-5021.

Bradburd GS, Ralph PL, Coop GM (2016) A spatial framework for understanding population structure and admixture. *PLoS Genetics* 12:e1005703.

Bradburd GS, Coop GM, Ralph PL (2018) Inferring continuous and discrete population genetic structure across space. *Genetics* 210:33-52.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23: 2633-2635.

Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL (2014) Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genet* 10:e1004410.

655 Brunsfeld S, Sullivan J, Soltis D, Soltis P (2001) Comparative
656 phylogeography of northwestern North America: a synthesis. *Special*
657 *Publication-British Ecological Society* 14:319-340.
658 Cardon LR, Palmer LJ (2003). Population stratification and spurious allelic
659 association. *The Lancet* 361(9357):598-604.
660 Charlesworth B, Charlesworth D (2010) *Elements of evolutionary genetics*.
661 Roberts and Company Publishers Greenwood Village, CO.
662 Corander J, Waldmann P, Marttinen P, Sillanpää MJ. 2004. BAPS 2:
663 enhanced possibilities for the analysis of genetic population structure.
664 *Bioinformatics* 20:2363-2369.
665 Danecek P, Auton A, Abecasis G, Albersm CA, Banks E, DePristo MA et
666 al. (2011) The variant call format and VCFtools. *Bioinformatics* 27(15):
667 2156–2158.
668 De Maio N, Schrempf D, Kosiol C (2015). PoMo: An Allele Frequency-
669 Based Approach for Species Tree Estimation. *Systematic Biology* 64:1018-
670 1031.
671 Dray S, Dufour AB (2007). The ade4 package: implementing the duality
672 diagram for ecologists. *Journal of Statistical Software*, 22(4), 1-20.
673 Ekblom R, Galindo J (2011). Applications of next generation sequencing in
674 molecular ecology of non-model organisms. *Heredity* 107(1):1–15.

675 Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES et al.
676 (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high
677 diversity species. *PLoS ONE* 6:e19379.
678 Falush D, Stephens M, Pritchard JK (2003). Inference of population
679 structure using multilocus genotype data: linked loci and correlated allele
680 frequencies. *Genetics* 164:1567-1587.
681 Frankham R (2010). Challenges and opportunities of genetic approaches to
682 biological conservation. *Biological Conservation* 143:1919-1927.
683 Friedman J, Twyford AD, Willis JH, Blackman BK (2015) The extent and
684 genetic basis of phenotypic divergence in life history traits in *Mimulus*
685 *guttatus*. *Molecular Ecology* 24:111-122.
686 Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshir RJ, Sun Q et al. 2014.
687 TASSEL-GBS: a high capacity genotyping by sequencing analysis
688 pipeline. *PloS one* 9: e90346.
689 Goudet J (2001) FSTAT, a program to estimate and test gene diversities and
690 fixation indices, version 2.9. 3. *http://www2.*
691 *unil.ch/popgen/softwares/fstat.htm*.
692 Hedin M, Starrett J, Hayashi C (2013). Crossing the uncrossable: novel
693 trans-valley biogeographic patterns revealed in the genetic history of low-
694 dispersal mygalomorph spiders (Antrodiaetidae, Antrodiaetus) from
695 California. *Molecular Ecology* 22:508-526.

696 Hendrick MF, Finseth FR, Mathiasson ME, Palmer KA, Broder EM,
697 Breigenzer P et al. (2016) The genetics of extreme microgeographic
698 adaptation: an integrated approach identifies a major gene underlying leaf
699 trichome divergence in Yellowstone *Mimulus guttatus*. *Molecular Ecology*
700 25,5647-5662.
701 House GL, Hahn MW (2018) Evaluating methods to visualize patterns of
702 genetic differentiation on a landscape. *Molecular Ecology Resources*
703 18:448-460.
704 Husband BC, Spencer CHB (1996) A metapopulation perspective in plant
705 population biology. *Journal of Ecology* 84:461-469.
706 Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009). Inferring weak
707 population structure with the assistance of sample group information.
708 *Molecular Ecology Resources* 9:1322-1332.
709 Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013)
710 diveRsity: an R package for the estimation and exploration of population
711 genetics parameters and their associated errors. *Methods in Ecology and*
712 *Evolution* 4:782-788.
713 Kooyers NJ, Greenlee AB, Colicchio JM, Oh M, Blackman BK (2015).
714 Replicate altitudinal clines reveal that evolutionary flexibility underlies
715 adaptation to drought stress in annual *Mimulus guttatus*. *New Phytologist*
716 206:152-165.

717 Lekberg Y, Roskilly B, Hendrick MF, Zabinski CA, Barr CM, Fishman L
718 (2012). Phenotypic and genetic differentiation among yellow monkeyflower
719 populations from thermal and non-thermal soils in Yellowstone National
720 Park. *Oecologia* 170:111-122.
721 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–
722 Wheeler transform. *Bioinformatics* 25:1754-1760.
723 Lindsay DW (1964) Natural dispersal of *Mimulus guttatus*. *Proc. Utah Acad.*
724 *Sci. Art. Lett.* 41:237-241.
725 Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of
726 a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American*
727 *Journal of Botany* 82:1420-1425.
728 Lowry DB, Rockwood RC, Willis JH (2008) Ecological reproductive
729 isolation of coast and inland races of *Mimulus guttatus*. *Evolution* 62:2196-
730 2214.
731 Lowry DB, Hall MC, Salt DE, Willis JH (2009). Genetic and physiological
732 basis of adaptive salt tolerance divergence between coastal and inland
733 *Mimulus guttatus*. *New Phytologist* 183(3): 776-788.
734 Lowry DB, Willis JH (2010). A widespread chromosomal inversion
735 polymorphism contributes to a major life-history transition, local adaptation,
736 and reproductive isolation. *PLOS Biology* 8: e1000500.

737 Lowry DB, Sobel JM, Angert AL, Ashman T-L, Baker RL, Blackman BK *et*
738 *al* (2019). The case for the continued use of the genus name *Mimulus* for all
739 monkeyflowers. *TAXON* **68**(4)**:** 617-623.
740 MacNair MR (1983). The genetic control of copper tolerance in the yellow
741 monkey flower, *Mimulus guttatus*. *Heredity* 50: 283-293.
742 Malecot G (1948) Les mathématiques de l'hérédité. Masson and Cie, Paris.
743 English translation. The mathematics of heredity. 1969. WH Freeman and
744 Co., San Francisco, CA. *Les mathématiques de l'hérédité. Masson and Cie,*
745 *Paris. English translation. The mathematics of heredity. 1969. WH Freeman*
746 *and Co., San Francisco, CA*.
747 Martin NH (2004) Flower size preferences of the honeybee (*Apis mellifera*)
748 foraging on *Mimulus guttatus* (Scrophulariaceae). Unpublished Thesis,
749 University of Texas https://digital.library.txstate.edu/handle/10877/2557.
750 McAssey EV, Corbi J, Burke JM (2016). Range-wide phenotypic and
751 genetic differentiation in wild sunflower. *BMC Plant Biology* 16: 249.
752 Meirmans PG (2012) The trouble with isolation by distance. *Molecular*
753 *Ecology* 21, 2839-2846.
754 Minh BQ, Nguyen MAT, von Haeseler A (2013) Ultrafast approximation for
755 phylogenetic bootstrap. *Molecular Biology and Evolution* 30:1188-1195.
756 Murray KD, Janes JK, Jones A, Bothwell HM, Andrew RL, Borevitz JO
757 (2019). Landscape drivers of genomic diversity and divergence in woodland
758 *Eucalyptus*. *Molecular Ecology* 28(24): 5232-5247.

759 Nathan R (2006) Long-distance dispersal of plants. *Science* 313, 786-788.
760 Nesom G (2014) Further observations on relationships in the *Erythranthe*
761 *guttata* group (Phrymaceae) *Phytoneuron* 93:1-8.
762 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A
763 fast and effective stochastic algorithm for estimating maximum-likelihood
764 phylogenies. *Molecular Biology and Evolution* 32:268-274.
765 Pedro T, Guillermo V-A, Carvalho SB (2015) Phylin: an R package for
766 phylogeographic interpolation. *Molecular Ecology Resources* 15:349-357.
767 Perez MF, Franco FF, Bombonato JR, Bonatelli IAS, Romeiro-Brito GKM,
768 Fegies AC et al. (2018) Assessing population structure in the face of isolation
769 by distance: Are we neglecting the problem? *Diversity and Distributions*
770 24:1883-1889.
771 Peterson ML, Kay KM, Angert AL (2016) The scale of local adaptation in
772 *Mimulus guttatus*: comparing life history races, ecotypes, and populations.
773 *New Phytologist* 211:345-356.
774 Petkova D, Novembre J, Stephens M. (2016) Visualizing spatial population
775 structure with estimated effective migration surfaces. *Nature Genetics*
776 48:94-100.
777 Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures
778 from genome-wide allele frequency data. *PLOS Genetics* 8:e1002967.
779 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population
780 structure using multilocus genotype data. *Genetics* 155:945-959.

781 Puzey JR, Willis JH, Kelly JK (2017) Population structure and local selection
782 yield high genomic variation in *Mimulus guttatus*. *Molecular Ecology*
783 26:519-535.
784 Raj A, Stephens M, Pritchard JK (2014). fastSTRUCTURE: variational
785 inference of population structure in large SNP data sets. *Genetics* 197:573-
786 589.
787 Reilly SB, Corl A, Wake DB (2015). An integrative approach to
788 phylogeography: investigating the effects of ancient seaways, climate, and
789 historical geology on multi-locus phylogeographic boundaries of the
790 Arboreal Salamander (*Aneides lugubris*). *BMC Evolutionary Biology*
791 15:241.
792 Ritland K, Ritland C (1996). Inferences about quantitative inheritance based
793 on natural population structure in the yellow monkeyflower, *Mimulus*
794 *guttatus*. *Evolution* 50:1074-1082.
795 Rousset F (1997) Genetic differentiation and estimation of gene flow from F-
796 statistics under isolation by distance. *Genetics* 145:1219-1228.
797 Rovito SM (2010). Lineage divergence and speciation in the Web-toed
798 Salamanders (Plethodontidae: Hydromantes) of the Sierra Nevada,
799 California. *Molecular Ecology* 19:4554-4571.
800 Schregel J, Remm J, Eiken HG, Swenson JE, Saarma U, Hagen SB (2018)
801 Multi-level patterns in population genetics: Variogram series detects a hidden

802 isolation-by-distance-dominated structure of Scandinavian brown bears
803 *Ursus arctos*. *Methods in Ecology and Evolution* 9:1324-1334.
804 Slatkin M (1987) Gene flow and the geographic structure of natural
805 populations. *Science* 236: 787-792.
806 Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium
807 populations. *Evolution* 47: 264-279.
808 Sweigart A, Karoly K, Jones A, Willis JH (1999) The distribution of
809 individual inbreeding coefficients and pairwise relatedness in a population of
810 *Mimulus guttatus*. *Heredity* 83:625-632.
811 Troth A, Puzey JR, Kim RS, Willis JH, Kelly JK (2018). Selective trade-offs
812 maintain alleles underpinning complex trait variation in plants. *Science*
813 361(6401): 475-478.
814 Truscott AM, Soulsby C, Palmer SCF, Newell L, Hulme PE (2006) The
815 dispersal characteristics of the invasive plant *Mimulus guttatus* and the
816 ecological significance of increased occurrence of high-flow events. *Journal*
817 *of Ecology* 94:1080-1091.
818 Twyford AD, Caola AM, Choudhary P, Raina R, Friedman J (2018). Loss of
819 color pigmentation is maintained at high frequency in a monkey flower
820 population. *American Naturalist* 191(1): 135-145.
821 Twyford AD, Friedman J (2015) Adaptive divergence in the monkey flower
822 *Mimulus guttatus* is maintained by a chromosomal inversion. *Evolution*
823 69:1476-1486.

824 Twyford AD, Streisfeld MA, Lowry DB, Friedman J (2015) Genomic studies
825 on the nature of species: adaptation and speciation in *Mimulus*. *Molecular*
826 *Ecology* 24:2601-2609.
827 Van Oosterhout C, Hutchinson WF, Wills DP, Shipley P (2004) MICRO-
828 CHECKER: software for identifying and correcting genotyping errors in
829 microsatellite data. *Molecular Ecology Notes* 4:535-538.
830 Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic
831 structure analyses in plant populations. *Molecular Ecology* 13:921-935.
832 Vickery RK (1999) Remarkable waxing, waning, and wandering of
833 populations of *Mimulus guttatus*: an unexpected example of global warming.
834 *The Great Basin Naturalist* 59:112-126.
835 Vickery RK, Phillips DR, Wonsavage PR (1986) Seed dispersal in *Mimulus*
836 *guttatus* by wind and deer. *The American Midland Naturalist* 116:206-208.
837 Waser NM, Vickery RK, Price MV (1982) Patterns of seed dispersal and
838 population differentiation in *Mimulus guttatus*. *Evolution* 36:753-761.
839 Whelan NV, Galaska MP, Sipley BN, Weber JM, Johnson PD, Halanych KM
840 et al. (2019). Riverscape genetic variation, migration patterns, and
841 morphological variation of the threatened Round Rocksnail, *Leptoxis ampla*.
842 *Molecular Ecology* 28(7): 1593-1610.
843 Wright S (1943) Isolation by distance. *Genetics* 28:114.

844 Wu CA, Lowry DB, Cooley AM, et al. (2007) *Mimulus* is an emerging model
845 system for the integration of ecological and genomic studies. *Heredity*
846 100:220-230.
847 Yuan Y-W, Rebocho AB, Sagawa JM, Stanley LE, Bradshaw HD (2016).
848 Competition between anthocyanin and flavonol biosynthesis produces spatial
849 pattern variation of floral pigments between *Mimulus* species. *Proceedings of*
850 *the National Academy of Sciences* 113(9): 2448-2453.

851 **Legends to figures**

852 **Figure 1.** Geographic locations and spatial scale of study of *Mimulus guttatus*
853 populations. (a) Fine-scale geographic sampling at two locations in
854 California. Five transects were sampled at different spacings in population
855 FOR (top panel) and ELD (bottom panel). Yellow dots indicate transects of
856 20 individual samples at 30cm spacing; inset in top panel highlights three
857 closely spaced transects, inset in bottom figure shows an example of the
858 detailed sampling of individuals performed for each transect, (b) population
859 sampling through the Sierra Nevada, (c) range-wide sampling, with the
860 known species limits marked with blue line. Note that a total of 81
861 populations were analysed in this study, with unique population samples for
862 each spatial scale.

863

864 **Figure 2.** Autocorrelograms showing the extent of spatial genetic structure
865 based on the kinship coefficient F*ij* as a function of distance for: (a) annual
866 population FOR, (b) perennial population ELD. Analyses are based on ten
867 PCR-based markers in population FOR and nine markers in ELD. Mean
868 values of F*ij* are shown for 11 distance classes per population. Faint lines
869 indicate 95% confidence intervals derived by permutation, and black bars are
870 standard errors derived by jack-knifing.

871

872 **Figure 3.** Genetic clustering of *M. guttatus* populations across the Sierra
873 Nevada. (a) Principal Component Analysis (PCA) showing individual
874 positions on PC1 and PC2. Individuals are coloured by their source
875 population to match the geographic map shown in inset, (b)
876 fastSTRUCTURE analysis with $K = 2$ reveals a north-south genetic divide,
877 (c) the best supported fastSTRUCTURE model $K = 6$ shows geographic
878 structure and population clustering. In (b) and (c), individuals are represented
879 by coloured bars, and assignment probability (Q-value) is displayed on the y-
880 axis.
881
882 **Figure 4.** Admixture bar plots for Sierra and range-wide *M. guttatus*
883 populations using conStruct. Each bar represents a population, which are
884 ordered by latitude (North-South), and assignment probability (Q-value) is
885 displayed on the y-axis. (a) conStruct non-spatial plot for 22,697 SNPs
886 present in 9 Sierra populations, (b) conStruct non-spatial plot using 3,414
887 SNPs present in range-wide populations, (c) conStruct spatial plot for Sierra
888 populations, (d) conStruct spatial plot for range-wide populations. Coloured
889 bar in panel (d) shows the presence of coastal populations in yellow and non-
890 coastal population in grey.
891
892 **Figure 5.** Maps of admixture proportions for *M. guttatus* conStruct spatial
893 and non-spatial analyses using *K*-values between 2 and 4. Pies show mean

admixture proportions across individuals from a given sampling site. Panels A-F show populations from the Sierra Nevada with colours to match Figure 3, with the scale bar representing 100 km. Panels G-L show range-wide populations, with green southern, orange coastal, purple northern, blue inland and teal central, with the scale bar representing 500 km.

**Figure 6.** Range-wide phylogeographic and diversity analyses of *M. guttatus* populations. (a) Map showing geographic variation in genetic diversity. Populations with more than two sampled individuals are coloured to indicate values for nucleotide diversity ($\pi$) per site per population, (b) Sampling map coloured by geographic clusters, modified from Twyford and Friedman (2015), (c) Maximum likelihood phylogeny generated in IQ-TREE using 72,941 SNPs and invariant sites scored in 70 populations, (d) TreeMix graph showing population splits inferred from 3,066 LD filtered SNPs present in the 30 populations with three sampled individuals. Ten migration edges shown, with migration weight indicated by the colour key.

**Figure 7.** Isolation by distance and genetic structure in *M. guttatus* across spatial scales. (a) Isolation by distance plot showing the association between geographic distance and genetic distance across the Sierra Nevada, (b) Isolation by distance plot for range-wide populations, (c) Semivariogram showing genetic divergence as a function of distance for Sierra Nevada

populations, with symbols showing semivariance for a given lag-distance and the line the best fitting model, (d) Semivariogram of genetic divergence relative to distance for range-wide populations.

**Figure S1.** Geographic map of samples used in the range-wide study. Sample population codes are coloured by geographic genetic clusters, with green southern, orange coastal, purple northern.

**Figure S2.** Cross-validation results for *M. guttatus* conStruct models. (a) Sierra data run with $K = 1$ through 9; (b) Range-wide data run with $K = 1$ through 10. Blue points represent the predictive accuracy for the spatial model, and green for the non-spatial model.

**Figure S3.** Percentage variance explained by TreeMix models of range-wide *M. guttatus* populations using varying levels of migration. Model fit for between 0 and 10 migration events assessed using the RADpipe package.

**Table S1.** Location information for newly sampled populations used in this study. Information is given for the sites used for studying fine-scale spatial genetic structure and population genetic variation. Details of the range-wide collections are reported in Twyford and Friedman (2015).