



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}

Received: 30 October 2019 / Revised: 2 June 2020 / Accepted: 15 June 2020 / Published online: 8 July 2020
© The Author(s), under exclusive licence to The Genetics Society 2020

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 recolonisation 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.

Associate editor: Marc Stift

Supplementary information The online version of this article (<https://doi.org/10.1038/s41437-020-0335-7>) contains supplementary material, which is available to authorized users.

✉ Alex D. Twyford
Alex.Twyford@ed.ac.uk

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, NY 13244, USA

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

⁵ Department of Biology, Queen's University, Kingston, ON K7L 3N6, Canada

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 recolonisation from glacial refugia can generate patterns that are hard to interpret (Slatkin 1987).

The development of new sequencing approaches make the study of natural populations more accessible (Ekblom

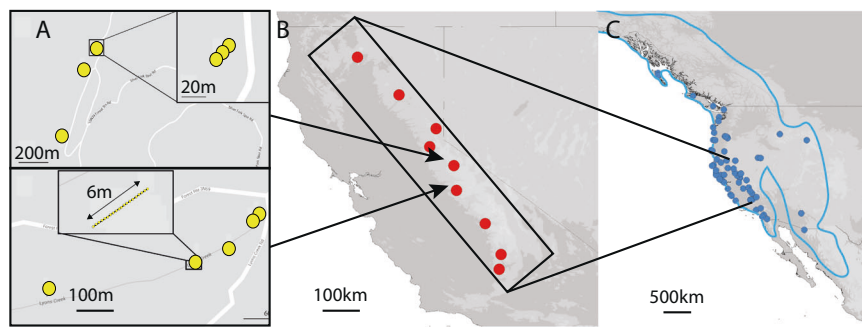


Fig. 1 Geographic locations and spatial scale of study of *Mimulus guttatus* populations. (a) Fine-scale geographic sampling at two locations in California. Five transects were sampled at different spacings in population FOR (top panel) and ELD (bottom panel). Yellow dots indicate transects of 20 individual samples at 30 cm spacing; inset in top panel highlights three closely spaced transects, inset in bottom

figure shows an example of the detailed sampling of individuals performed for each transect, (b) population sampling through the Sierra Nevada, (c) range-wide sampling, with the known species limits marked with blue line. Note that a total of 81 populations were analysed in this study, with unique population samples for each spatial scale.

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 be challenging against the background of continuous genetic variation often present in nature. As such, group assignment algorithms may detect artificial genetic clusters in populations characterised by isolation by distance, especially if the sampling of populations is aggregated (Pritchard et al. 2000). Nevertheless, a recent meta-analysis has shown that isolation by distance is present in 60% of population genetic datasets, with more than half these studies continuing to use the programme STRUCTURE without accounting for geographic distances between sampling locations (Perez et al. 2018). The long-running issue of identifying ‘clusters versus clines’ motivated the development of new methods that account for spatial information of samples when estimating genetic structure (Bradburd et al. 2016, 2018; Petkova et al. 2016; House and Hahn 2018), and these methods may be more accurate in detecting distinct genetic clusters in the presence of continuous patterns of genetic differentiation. However, only a few studies to date have applied these methods (e.g., Murray et al. 2019; Whelan et al. 2019).

A further issue with studying genetic structure is going from pattern to process, when a range of scale-dependent and temporally variable processes together shape the spatial arrangement of genetic variation (Schregel et al. 2018). At a local scale, the amount of dispersal and the strength and pattern of microgeographic selection, are expected to jointly affect spatial dynamics and patterns of relatedness (Vekemans and Hardy 2004). At the population level, barriers to dispersal and regional selection will determine

metapopulation structure and consequently evolutionary dynamics (Husband and 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 population genetic divergence in this species is high ($F_{ST} \sim 0.5$; Lowry et al. 2008; Puzey et al. 2017), therefore we may expect highly structured populations, with localised genetic clusters and limited gene flow. In contrast, owing to its high dispersal potential through seeds and vegetative fragments (Lindsay 1964; Truscott et al. 2006; Vickery et al. 1986; Waser et al. 1982), *M. guttatus* has spread to the farthest reaches of the Aleutian Islands in Alaska and rapidly invaded large areas of north-western Europe and New Zealand over the past c.

200 years (Truscott et al. 2006). Based on these observations, one would predict broad-scale genetic structure but limited structure at a fine spatial scale.

We address these different expectations by investigating genetic variation and population structure at a range of spatial scales (Fig. 1). First, we genotype individuals at a fine-spatial scale to look for spatial genetic structure (SGS) and limits to localised dispersal within two *M. guttatus* populations. Second, we use reduced-representation sequencing of populations across a 700 km transect from the Sierra Nevada to infer the nature of genetic structure, using conStruct (Bradburd et al. 2018), a recently developed method to infer discrete genetic clusters from continuous population samples. Third, we analyse sequence variation from range-wide populations to infer the extent of divergence and the distribution of genetic variation, and to understand the 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 horizontal spreading stems (stolons), which may facilitate local clonal spread and may also break-off and disperse along watercourses.

The extensive distribution range of *M. guttatus*, with a native range extending over 5000 km from northern Mexico

to Alaska, makes it a useful study system for investigating geographic genetic structure and responses to biogeographic barriers. *M. guttatus* is widespread and abundant in areas with a rich biota and complex biogeography, encircling the Central Valley, spanning the Cascade/Sierran transition, bridging the Cascades/Coast ranges and the Rocky Mountains, and found in formerly glaciated regions of western Canada and Alaska. Previous studies of *M. guttatus* populations have identified geographic genetic structure corresponding to coastal and inland populations (Lowry et al. 2008) or northern, coastal, and southern populations (Twyford and Friedman 2015). However, the confounding issue of isolation by distance observed in some population studies of *M. guttatus* (e.g., Kooyers et al. 2015), and the lack of support for some nodes in phylogeographic analyses (Twyford and Friedman 2015), has precluded detailed interpretation of geographic genetic structure and the phylogeographic history of the species.

Fine-scale spatial genetic structure

We used two populations from California to estimate spatial genetic structure. 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 and 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 sampling distances between transects 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 Analyser 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 alleles and other errors with MICRO-CHECKER (Van Oosterhout et al. 2004). We excluded locus MgSTS278 in population ELD due to uneven amplification success, prior to statistical analyses. The final dataset included 91 individuals for population ELD and 79 for FOR.

We inferred individuals likely to be the product of clonal reproduction using a permutation and re-sampling approach that accounts for scoring error and somatic mutations, as implemented in GENCLONE (Arnaud-Haond and Belkhir 2007). We related clonality to the inter-plant sampling distance to understand the extent of clonal spread, then selected a single individual at random from each clone for downstream analyses of diversity and relatedness. We calculated the extent of SGS for each population using spatial autocorrelation analysis described in Vekemans and Hardy (2004), using the pairwise kinship coefficients (F_{ij}) of Loiselle et al. (1995). We performed analyses with SPA-GeDi (Vekemans and Hardy 2004) using the following distance classes: 0–2 m, 2–4 m, 4–6 m, 6–8 m, 8–10 m, 10–20 m, 20–50 m, 50–100 m, 100–200 m, 200–400 m and 400–700 m. We calculated mean F_{ij} per distance class, 95% confidence intervals by permutation, standard errors by jack-knifing, and plotted autocorrelograms for each analysis. We calculated overall spatial genetic structure per population with the S_p statistic. As we found little evidence of genetic 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 95 km 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 adaptors, 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 and 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, yielding less than 1% of the mean number of sequencing reads across samples, while other samples yielded between 766,539–6,704,555 reads. The average sequencing coverage per site was 38-fold, for 5,611,458 sites. Downstream population genetic analyses used a subset of data filtered to include individuals with less than 50% missing data, sites scored in over 75% individuals, a minor allele frequency of 0.05, and with SNPs in tight linkage removed by filtering variants within 20 bp (Brandvain et al. 2014), to give a final dataset of 22,697 SNPs.

We inferred discrete population structure using conStruct (Bradburd et al. 2018), which models admixture across a specified number of discrete layers as defined by the K -value. Non-spatial conStruct analyses do not use location information, while spatial conStruct analyses assume allele frequencies have a positive covariance based on geographic locations to account for isolation by distance. To determine an appropriate level of parameterisation for the models, we used cross-validation with a training set (Bradburd et al. 2018), and compared predictive accuracies between spatial and non-spatial models, and between successive K -values, to determine which model has the best goodness-of-fit without overfitting. We analysed K -values of 1–9. To test whether spatial models were better fitting than non-spatial models we used paired t -tests comparing cross-validation scores across values of K . The best 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 maximises 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 (π) 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

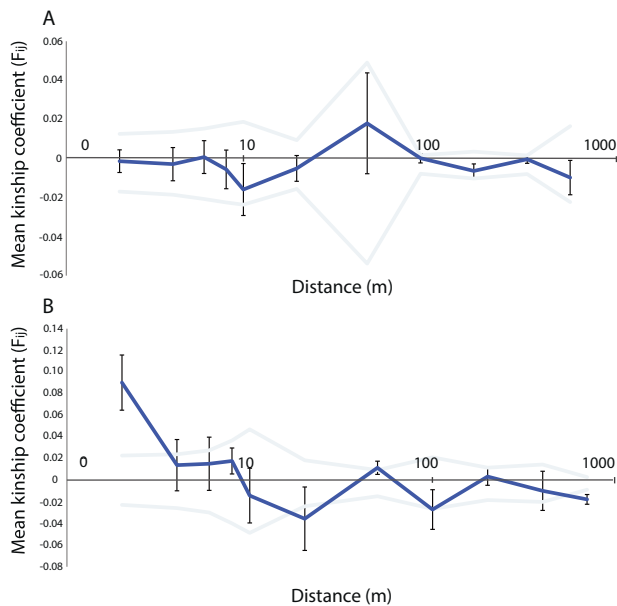


Fig. 2 Autocorrelograms showing the extent of spatial genetic structure based on the kinship coefficient F_{ij} as a function of distance for. (a) annual population FOR, (b) perennial population ELD. Analyses are based on ten PCR-based markers in population FOR and nine markers in ELD. Mean values of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

outside a chromosomal inversion by Twyford and 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) (Fig. S1). We recalled 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 6523 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 than 25% missing data, producing a dataset of 3414 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 3066 filtered SNPs.

We used conStruct, as described above (but with K -values between 1 and 10), to characterise 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 population. 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 and 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 π 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 π 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, Dufour 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.

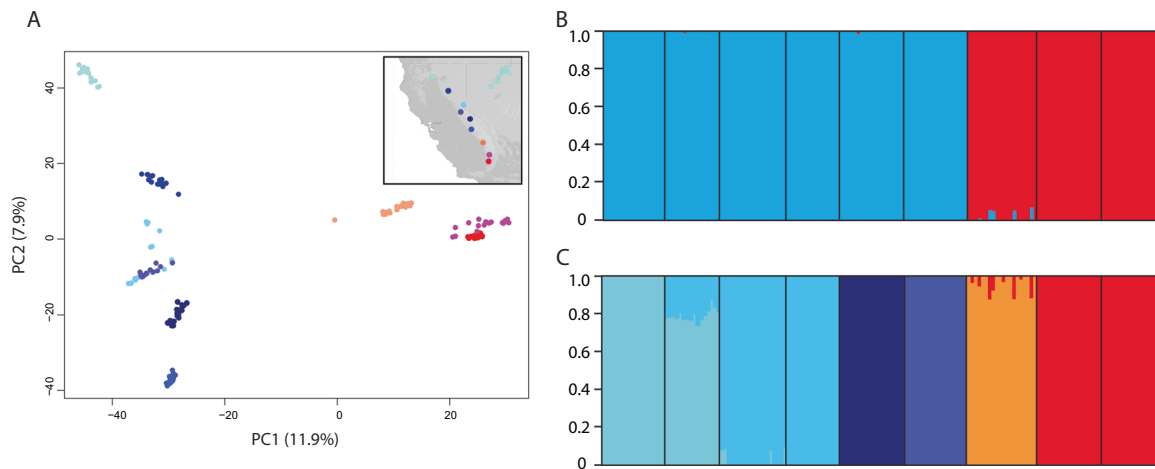
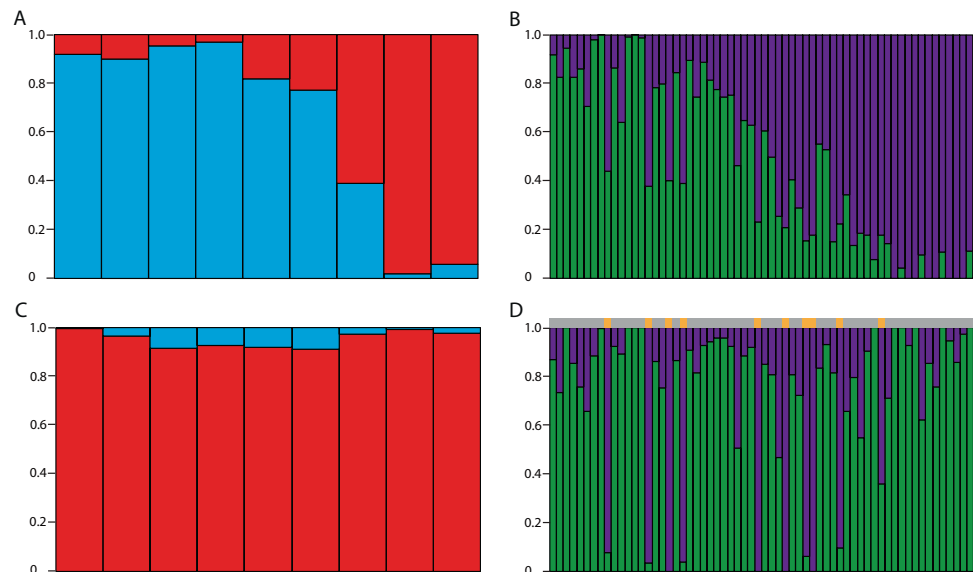


Fig. 3 Genetic clustering of *M. guttatus* populations across the Sierra Nevada. (a) Principal Component Analysis (PCA) showing individual positions on PC1 and PC2. Individuals are coloured by their source population to match the geographic map shown in inset, (b) fastSTRUCTURE analysis with $K = 2$ reveals a north-south genetic

divide, (c) the best supported fastSTRUCTURE model $K = 6$ shows geographic structure and population clustering. In **b** and **c**, individuals are represented by coloured bars, and assignment probability (Q-value) is displayed on the y-axis.

Fig. 4 Admixture bar plots for Sierra and range-wide *M. guttatus* populations using conStruct. Each bar represents a population, which are ordered by latitude (North-South), and assignment probability (Q-value) is displayed on the y-axis. (a) conStruct non-spatial plot for 22,697 SNPs present in 9 Sierra populations, (b) conStruct non-spatial plot using 3414 SNPs present in range-wide populations, (c) conStruct spatial plot for Sierra populations, (d) conStruct spatial plot for range-wide populations. Coloured bar in panel **d** shows coastal populations in yellow and non-coastal population in grey.



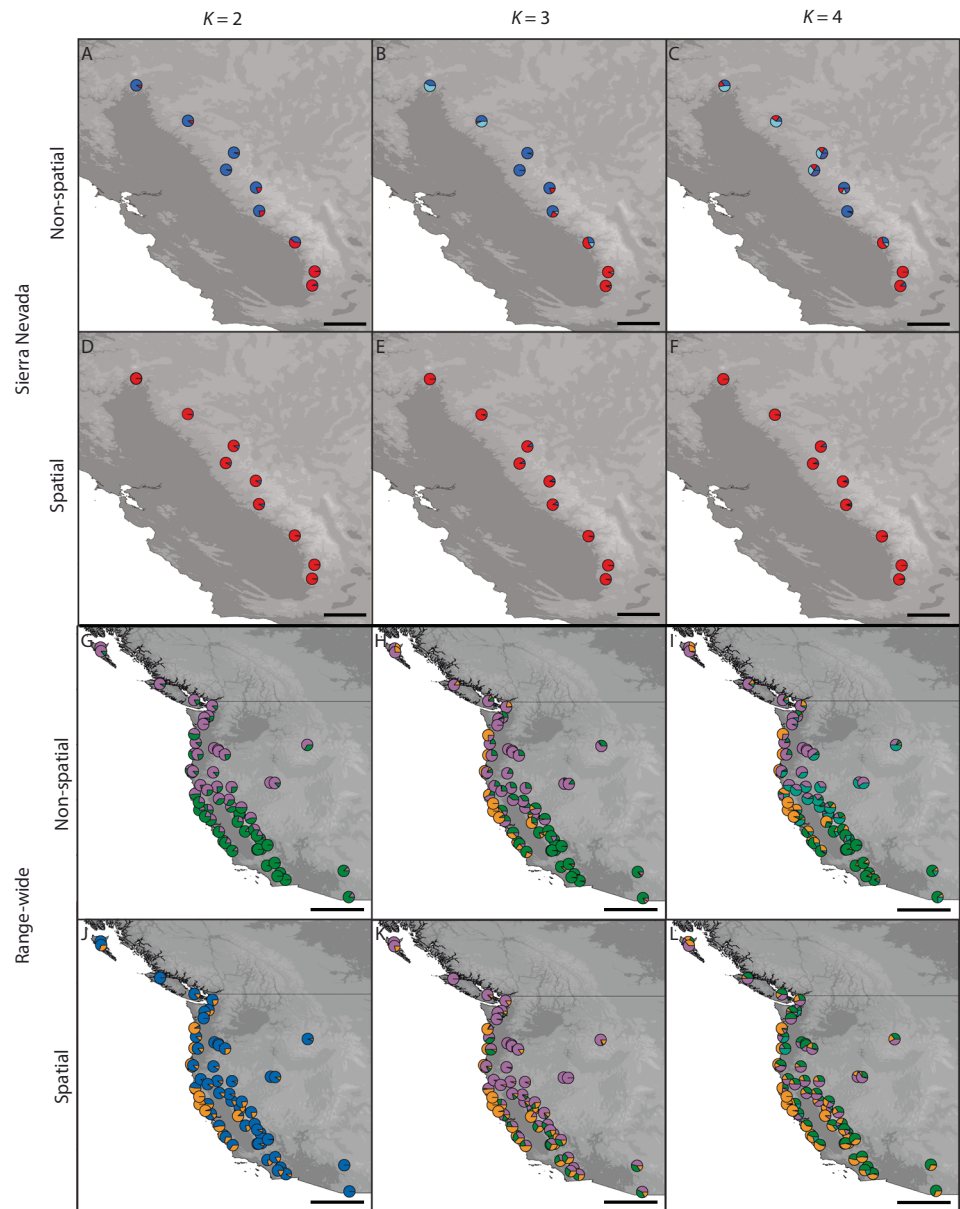
Results

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 ($S_p = -5.90 \times 10^{-5}$). Over all distance classes, values of F_{ij} consistently fell within the permuted upper and lower confidence intervals, reflecting no spatial structure (Fig. 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 (Fig. 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 metres 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 π of 1.6%, and high population structure with a mean pairwise F_{ST} of 0.327. Analyses of genetic clustering using

Fig. 5 Maps of admixture proportions for *M. guttatus* conStruct spatial 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 Fig. 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.



PCA and fastSTRUCTURE revealed geographic genetic clusters corresponding to northern and southern Sierra Nevada populations (Fig. 3). Similarly, the non-spatial model in conStruct showed a north-south genetic division at $K = 2$ (Fig. 4a), with genetic clusters corresponding to geographic groupings at $K = 3$ or 4 (Fig. 5a–c; with $K = 4$ the value at which the likelihood plateaus, Fig. 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 (Fig. S2). The spatial conStruct models did not show a clear north-south genetic discontinuity at $K = 2$ (Fig. 4c), and at higher K -values (Fig. 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 ($R^2 = 0.545$; Figs. 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 ($P < 0.0001$, Fig. S2), revealed no such correlation with latitude at $K = 2$ ($R^2 = 0.0001$), and instead discriminated coastal from inland populations (Fig. 5j). At $K = 3$, spatial models distinguished coastal, northern and southern genetic clusters but with major admixture (Fig. 5k).

Complex patterns of genetic structure were also evident in polymorphism aware phylogenetic analyses, which resolved a well-supported tree topology with genetic

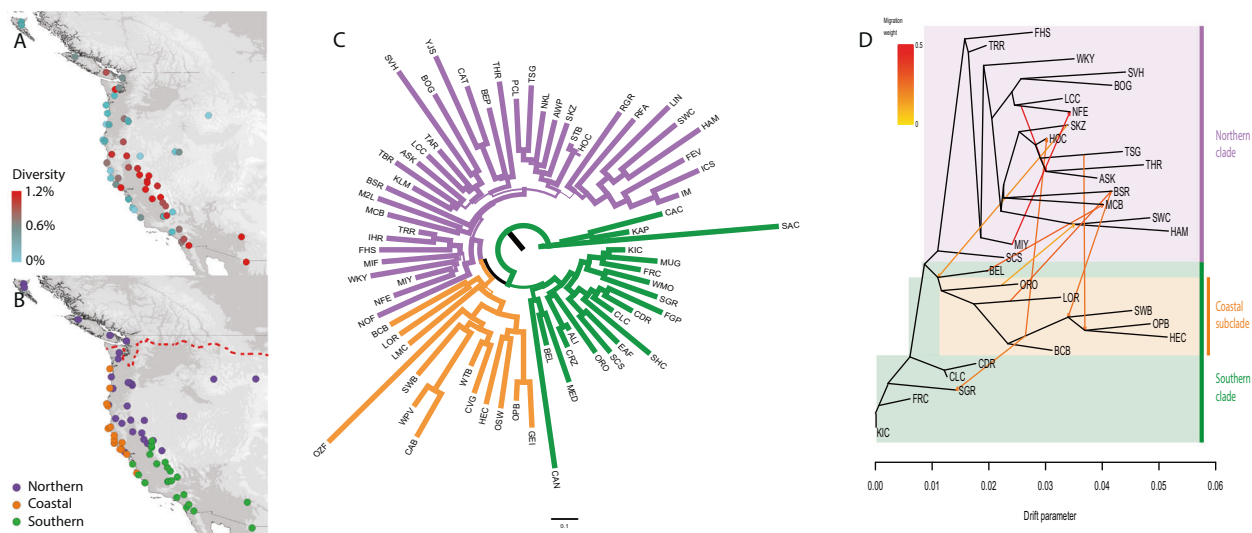


Fig. 6 Range-wide phylogeographic and diversity analyses of *M. guttatus* populations. (a) Map showing geographic variation in genetic diversity. Populations with more than two sampled individuals are coloured to indicate values for nucleotide diversity (π) per site per population, (b) Sampling map coloured 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 3066 LD filtered SNPs present in the 30 populations with three sampled individuals. Ten migration edges shown, with migration weight indicated by the colour key.

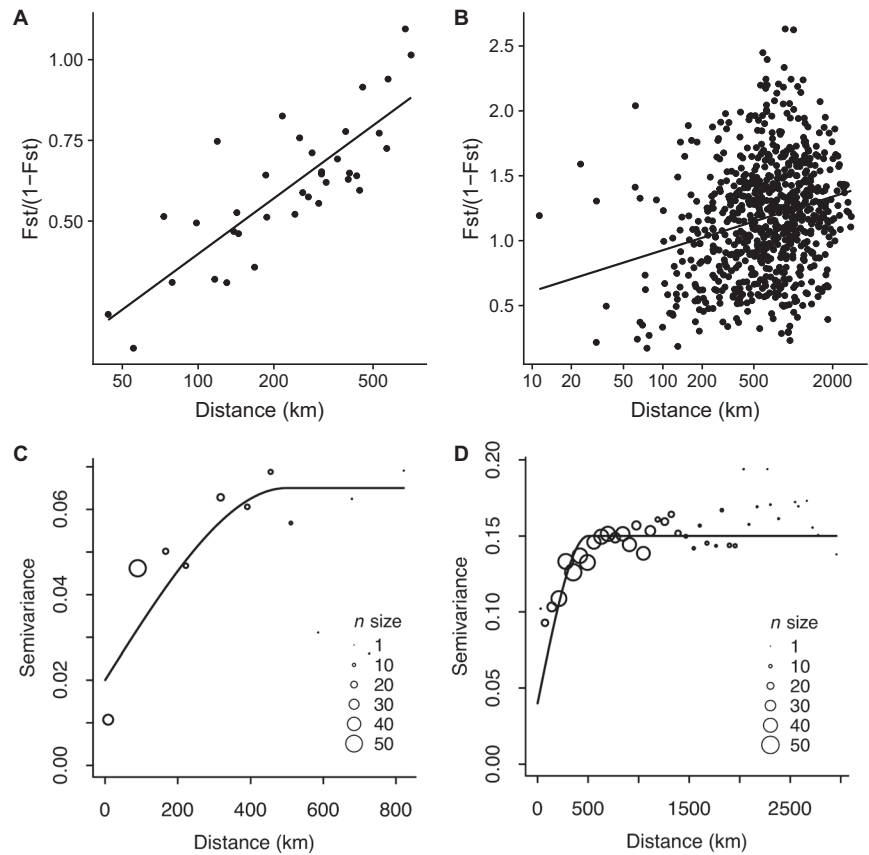
clusters of southern, coastal and northern populations, while also revealing previously uncharacterised geographic substructure within clades (Fig. 6). For example, well-supported at the base of the coastal clade are two Californian populations from Monterey Bay, BCB and LOR, with other more northerly 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 (Fig. S3). At $m = 10$, dispersal is observed across the admixture graph, including multiple dispersal events from populations in California (Fig. 6c). These results indicate a history of repeated dispersal across the range of *M. guttatus*, facilitating recolonisation after glaciation.

Genetic diversity as estimated by population-level π 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 (Fig. 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) and in the more sparsely sampled inland southern populations of Sonora (ALI) and Arizona (CRZ).

For the Sierra dataset, the linear regression between $(F_{ST} / 1 - F_{ST})$ and log pairwise geographic distance was significant ($P < 0.001$) and geographic distance explained 66% of genetic variation (Fig. 7a). However, for the range-wide dataset, although the linear regression is significant ($P < 0.001$), the data showed a poor fit, with distance explaining only 6% of genetic variation (Fig. 7b). This matches predictions that isolation by distance should break down with increasing geographic distance as dispersal processes change and as different geographic genetic clusters mix. Similarly, Mantel tests for the correlation between the geographic and genetic distance matrices showed a significant correlation for the Sierra dataset ($r: 0.69$, $P < 0.01$), but no significant correlation for the range-wide dataset ($r: 0.13$, $P = 0.13$). To provide a separate estimate of the geographic scale of genetic differentiation independent of mutation rate we related genetic divergence to geographic distance in variogram models. Our results showed that the range, defined as the scale of spatial autocorrelation after which little change in the semi-variance is encountered with increasing distances, extended to 500 km for both Sierra and range-wide GBS data (Fig. 7c, d). These analyses showed that the spatial independence of populations is only achieved at 500 km.

Fig. 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.



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 this species (Peterson et al. 2016; Troth et al. 2018). We discuss our results below in terms of the spatial scale of gene flow and the species' historical demography, and make recommendations for how best to use genetic information to infer genetic structure at different spatial scales.

Spatial dynamics of dispersal and migration

Our findings show that high dispersal potential has shaped genetic structure of *M. guttatus* populations. The lack of spatial genetic structure over hundreds of metres likely reflects extensive local dispersal. Similarly, the emergence of broad geographic genetic clusters that extend over potential barriers to dispersal, and variogram analyses that reveal the non-independence of populations over hundreds of kilometres, suggests large metapopulations. Although pollen movement by bees, and downstream dispersal of vegetative fragments will contribute to dispersal in this species, seed-mediated dispersal is likely to dominate. While >40% of *M. guttatus* seeds fall within 25 cm of the

maternal plant (Ritland and Ritland 1996; Sweigart et al. 1999; Vickery et al. 1986), giving rise to some localised fine-scale genetic structure (Ritland and Ritland 1996) as observed in the perennial population, many of the light-weight seeds (0.002 mg) are likely dispersed much further. Occasional long-distance seed dispersal by wind, animals or water (Martin 2004) may be crucial for the widespread colonisation of *Mimulus* in its native range, and also in its introduced range where it has become a dominant species of disturbed watercourses over the last 200 years. Taken together, our analyses and previous work show the important role of dispersal at all spatial scales.

Research on genetic population structure in natural populations is often focussed on assessing genetic divergence (e.g., $F_{ST} \sim 1/(4N_e m + 1)$) and understating demographic connectivity by predicting migration rate (m). However, linking F_{ST} and m relies on a number of assumptions that may be unrealistic in most natural populations. For example, the challenge for understanding migration is illustrated by the contrast between high F_{ST} values in *M. guttatus* which imply low migration (this study, and others reviewed in Puzey et al. 2017), and a migration rate sufficient to homogenise population differences in models fit to whole genome data (Aeschbacher et al. 2017). Crucial to estimating migration is understanding diversity and the effective population size (N_e), with previous estimates of N_e for *M. guttatus* in the hundreds of thousands (between 4.805×10^5 and 6.730×10^5 : Aeschbacher et al. 2017; Brandvain et al. 2014), while the synonymous π value of 3.3% estimated by Puzey et al. (2017) makes *M. guttatus* one of the most genetically diverse plant species studied to date. Our study supports the finding of high genetic variation maintained in *M. guttatus* populations—we found up to thirty alleles at polymorphic markers within a population, numerous unique genotypes in a clonal perennial population, and a high value for sequence diversity at 1.6%. High genetic diversity was present even in populations with moderate to high selfing rates. Nonetheless, the demography of *M. guttatus* populations are characterised by ‘boom and bust’ dynamics, with rapid colonisation and population expansion in response to ecological opportunities, but with frequent local extinctions due to drought and habitat change (Vickery 1999). Overall, it is possible that while seed dispersal allows substantial mixing within populations and facilitates occasional long-distance dispersal, migration between populations is not always sufficient to homogenise population differences.

Spatial scale of genetic structure

Our work highlights the confounding influence of continuous genetic variation on the inference of genetic clusters. At regional spatial scales, for example across Sierra

Nevada populations of *M. guttatus*, we found strong isolation by distance, and spatial analyses accounting for continuous population structure did not detect clear geographic genetic structure. This indicates continuous genetic variation with geographic clines in allele frequencies, rather than discrete population clusters due to barriers to dispersal. Nonetheless, there was some evidence of subtle north-south divergence in the Sierra Nevada, which is notably less distinct than in other organisms. In spiders, for example, cryptic intraspecific breaks and species divergence were found between Sierra populations (Hedin et al. 2013). 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 North America. 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 Nevada. Whether there was a separate coastal refugium is hard to say with certainty. A coastal refugium is recognised for many North American plant species (Brunsfeld et al. 2001), and the patchy occurrence of high genetic diversity in coastal populations, and the topology of the phylogeny, are broadly consistent with a separate coastal refugium at the southern extent of the Northwest Forested Mountain biogeographic area, in the region of the Wilson Grove Formation. However, evidence for shared genetic variants between coastal populations and a population approximately 200 km inland, East of the Central Valley (population MED), suggests coastal populations may be independently derived from an

inland source, rather than from a separate coastal refugium. Similar patterns of trans-valley relatedness have been seen in spiders (Hedin et al. 2013) and salamanders (Reilly et al. 2015). Major rivers are orientated in a perpendicular axis to the Sierra Nevada mountains (Rovito 2010), and may have acted as a route for dispersal of the perennial ecotype. Our results support the model of Western North American phylogeography proposed by Brunsfeld et al. (2001), where vicariance, dispersal and refugia shape genomic variation, and where dispersal has occurred in waves as postglacial conditions became more hospitable.

Subsequent range expansion from glacial refugia has left a clear genetic signature, with a latitudinal cline of genetic variation across the north of the species range that parallels broad-scale north-south divergence seen in other taxa such as wild sunflowers (McAssey et al. 2016). Northwards range expansion and a subsequent increase in population size is likely to be recent, within the last ~20,000 years (Brandvain et al. 2014), as a response to increased habitat availability and more hospitable conditions post-glaciation.

The improved branch support in polymorphism-aware phylogenetic analyses compared with conventional Bayesian analysis of concatenated sequences (Twyford and Friedman 2015), supports the divergence of northern and coastal populations, and suggests a scenario of multiple independent colonisation events from the south, each with different biogeographic histories. Of particular interest is the coastal genetic cluster, which is mostly restricted to a narrow band adjacent to the Pacific. The low genetic diversity suggests these populations have been through a genetic bottleneck, while the TreeMix analysis suggests these populations have subsequently been a major source of admixture with inland populations. Overall, range-wide genetic variation in *M. guttatus* has been shaped by recurrent colonisation from the south of the species range, with dispersal avenues facilitating colonisation.

Our finding of multiple independent recolonisation events has important consequences for selecting samples for demographic analyses of *M. guttatus*. Genomic studies using sparse population samples need to compare individuals derived from a similar range expansion event, otherwise demographic inferences will reflect ancestral variation rather than recent population changes. While our data support (at least) two colonisation events from the south of the species range, recolonisation from a northern refugium, such as the Berengian refugia proposed for cold-tolerant taxa such as the serrated wintergreen *Orthilia secunda* (Beatty and 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 as make use of spatial analyses of genetic structure to better understand genetic variation in widespread species.

Data availability

The marker genotype data and SNP matrix are available in Dryad (<https://doi.org/10.5061/dryad.ngf1vhr8>).

Acknowledgements We thank Simon Aeschbacher for useful discussions of models of gene flow, Gideon Bradburd for advice on conStruct, Hannes Becher for help with R code, Jaanus Remm for assistance with interpreting variograms and Mabon Elis for help with figures. This research was supported by a Natural Environment Research Council Fellowship (NE/L011336/1) and a Heredity Fieldwork Grant to ADT, and a National Science Foundation grant (DEB-1354259) to JF. The fine-scale spatial genetics component formed part of the MSc thesis of ELYW at the RBGE, which is funded from the Scottish Government's Rural and Environment Science and Analytical Services Division (RESAS).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Aeschbacher S, Selby JP, Willis JH, Coop G (2017) Population-genomic inference of the strength and timing of selection against gene flow. *Proc Natl Acad Sci USA* 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. *Mol Ecol 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*. *Mol Ecol* 19:5009–5021
- Bradburd GS, Coop GM, Ralph PL (2018) Inferring continuous and discrete population genetic structure across space. *Genetics* 210:33–52
- Bradburd GS, Ralph PL, Coop GM (2016) A spatial framework for understanding population structure and admixture. *PLoS Genet* 12:e1005703
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckley 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
- Brunsfeld S, Sullivan J, Soltis D, Soltis P (2001) Comparative phylogeography of northwestern North America: a synthesis. *Spec Publ-Br Ecol Soc* 14:319–340
- Danecek P, Auton A, Abecasis G, Albersm CA, Banks E, DePristo MA et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158
- De Maio N, Schrepf D, Kosiol C (2015) PoMo: an allele frequency-based approach for species tree estimation. *Syst Biol* 64:1018–1031
- Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw* 22(4):1–20.
- Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 107(1):1–15
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckley ES et al. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587
- Friedman J, Twyford AD, Willis JH, Blackman BK (2015) The extent and genetic basis of phenotypic divergence in life history traits in *Mimulus guttatus*. *Mol Ecol* 24:111–122
- Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshir RJ, Sun Q et al. (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* 9:e90346
- Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices, version 2.9. 3. <http://www2.unil.ch/popgen/softwares/fstat.htm>
- Hedin M, Starrett J, Hayashi C (2013) Crossing the uncrossable: novel trans-valley biogeographic patterns revealed in the genetic history of low-dispersal mygalomorph spiders (Antrodiaetidae, Antrodiaetus) from California. *Mol Ecol* 22:508–526
- Hendrick MF, Finseth FR, Mathiasson ME, Palmer KA, Broder EM, Breigenzer P et al. (2016) The genetics of extreme microgeographic adaptation: an integrated approach identifies a major gene underlying leaf trichome divergence in Yellowstone *Mimulus guttatus*. *Mol Ecol* 25:5647–5662
- House GL, Hahn MW (2018) Evaluating methods to visualize patterns of genetic differentiation on a landscape. *Mol Ecol Resour* 18:448–460
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* 9:1322–1332
- Husband BC, Spencer CHB (1996) A metapopulation perspective in plant population biology. *J Ecol* 84:461–469
- Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol Evol* 4:782–788
- Kooyers NJ, Greenlee AB, Colicchio JM, Oh M, Blackman BK (2015) Replicate altitudinal clines reveal that evolutionary flexibility underlies adaptation to drought stress in annual *Mimulus guttatus*. *N. Phytol* 206:152–165
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760
- Lindsay DW (1964) Natural dispersal of *Mimulus guttatus*. *Proc Utah Acad Sci Art Lett* 41:237–241
- Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am J Bot* 82:1420–1425
- Lowry DB, Hall MC, Salt DE, Willis JH (2009) Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland *Mimulus guttatus*. *N Phytologist* 183(3):776–788
- Lowry DB, Rockwood RC, Willis JH (2008) Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. *Evolution* 62:2196–2214
- Lowry DB, Sobel JM, Angert AL, Ashman T-L, Baker RL, Blackman BK et al. (2019) The case for the continued use of the genus name *Mimulus* for all monkeyflowers. *Taxon* 68(4):617–623
- Lowry DB, Willis JH (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol* 8:e1000500
- Malecot G (1948) Les mathématiques de l'hérédité. Masson and Cie, Paris. English translation. The mathematics of heredity. 1969. WH Freeman and Co., San Francisco, CA. *Les mathématiques de l'hérédité. Masson and Cie, Paris. English translation. The mathematics of heredity. 1969. WH Freeman and Co., San Francisco.*
- Martin NH (2004) Flower size preferences of the honeybee (*Apis mellifera*) foraging on *Mimulus guttatus* (Scrophulariaceae). Unpublished Thesis, University of Texas
- McAssey EV, Corbi J, Burke JM (2016) Range-wide phenotypic and genetic differentiation in wild sunflower. *BMC Plant Biol* 16:249
- Minh BQ, Nguyen MAT, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 30:1188–1195
- Murray KD, Janes JK, Jones A, Bothwell HM, Andrew RL, Borevitz JO (2019) Landscape drivers of genomic diversity and divergence in woodland *Eucalyptus*. *Mol Ecol* 28(24):5232–5247
- Nathan R (2006) Long-distance dispersal of plants. *Science* 313:786–788
- Nesom G (2014) Further observations on relationships in the *Erythranthe guttata* group (Phrymaceae). *Phytoneuron* 93:1–8
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274
- Pedro T, Guillermo V-A, Carvalho SB (2015) Phylin: an R package for phylogeographic interpolation. *Mol Ecol Resour* 15:349–357
- Perez MF, Franco FF, Bombonato JR, Bonatelli IAS, Romeiro-Brito GKM, Fegies AC et al. (2018) Assessing population structure in the face of isolation by distance: are we neglecting the problem? *Diver Distrib* 24:1883–1889
- Peterson ML, Kay KM, Angert AL (2016) The scale of local adaptation in *Mimulus guttatus*: comparing life history races, ecotypes, and populations. *N Phytol* 211:345–356
- Petkova D, Novembre J, Stephens M (2016) Visualizing spatial population structure with estimated effective migration surfaces. *Nat Genet* 48:94–100
- Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8:e1002967
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Puzey JR, Willis JH, Kelly JK (2017) Population structure and local selection yield high genomic variation in *Mimulus guttatus*. *Mol Ecol* 26:519–535
- Reilly SB, Corl A, Wake DB (2015) An integrative approach to phylogeography: investigating the effects of ancient seaways, climate, and historical geology on multi-locus phylogeographic boundaries of the arboreal Salamander (*Aneides lugubris*). *BMC Evol Biol* 15:241
- Ritland K, Ritland C (1996) Inferences about quantitative inheritance based on natural population structure in the yellow monkey-flower, *Mimulus guttatus*. *Evolution* 50:1074–1082

- Rovito SM (2010) Lineage divergence and speciation in the Web-toed Salamanders (Plethodontidae: Hydromantes) of the Sierra Nevada, California. *Mol Ecol* 19:4554–4571
- Schregel J, Remm J, Eiken HG, Swenson JE, Saarma U, Hagen SB (2018) Multi-level patterns in population genetics: Variogram series detects a hidden isolation-by-distance-dominated structure of Scandinavian brown bears *Ursus arctos*. *Methods Ecol Evolution* 9:1324–1334
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* 236:787–792
- Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47:264–279
- Sweigart A, Karoly K, Jones A, Willis JH (1999) The distribution of individual inbreeding coefficients and pairwise relatedness in a population of *Mimulus guttatus*. *Heredity* 83:625–632
- Troth A, Puzey JR, Kim RS, Willis JH, Kelly JK (2018) Selective trade-offs maintain alleles underpinning complex trait variation in plants. *Science* 361(6401):475–478
- Truscott AM, Soulsby C, Palmer SCF, Newell L, Hulme PE (2006) The dispersal characteristics of the invasive plant *Mimulus guttatus* and the ecological significance of increased occurrence of high-flow events. *J Ecol* 94:1080–1091
- Twyford AD, Caola AM, Choudhary P, Raina R, Friedman J (2018) Loss of color pigmentation is maintained at high frequency in a monkey flower population. *Am Naturalist* 191(1):135–145
- Twyford AD, Friedman J (2015) Adaptive divergence in the monkey flower *Mimulus guttatus* is maintained by a chromosomal inversion. *Evolution* 69:1476–1486
- Twyford AD, Streisfeld MA, Lowry DB, Friedman J (2015) Genomic studies on the nature of species: adaptation and speciation in *Mimulus*. *Mol Ecol* 24:2601–2609
- Van Oosterhout C, Hutchinson WF, Wills DP, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4:535–538
- Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol Ecol* 13:921–935
- Vickery RK (1999) Remarkable waxing, waning, and wandering of populations of *Mimulus guttatus*: an unexpected example of global warming. *Gt Basin Naturalist* 59:112–126
- Vickery RK, Phillips DR, Wonsavage PR (1986) Seed dispersal in *Mimulus guttatus* by wind and deer. *Am Midl Naturalist* 116:206–208
- Waser NM, Vickery RK, Price MV (1982) Patterns of seed dispersal and population differentiation in *Mimulus guttatus*. *Evolution* 36:753–761
- Whelan NV, Galaska MP, Siple BN, Weber JM, Johnson PD, Halanych KM et al. (2019) Riverscape genetic variation, migration patterns, and morphological variation of the threatened Round Rocksnail, *Leptoxis ampla*. *Mol Ecol* 28(7):1593–1610
- Wright S (1943) Isolation by distance. *Genetics* 28:114
- Wu CA, Lowry DB, Cooley AM et al. (2007) *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity* 100:220–230
- Yuan Y-W, Rebocho AB, Sagawa JM, Stanley LE, Bradshaw HD (2016) Competition between anthocyanin and flavonol biosynthesis produces spatial pattern variation of floral pigments between *Mimulus* species. *Proc Natl Acad Sci USA* 113(9):2448–2453