Received: 19 March 2024 | Accepted: 1 July 2024

DOI: 10.1002/ajb2.16413

RESEARCH ARTICLE

Botany

Population structure and natural selection across a flower color polymorphism in the desert plant *Encelia farinosa*

Sonal Singhal[1] | Christopher DiVittorio[2,3] | Chandra Jones[1] | Itzel Ixta[1] | Alexis Widmann[1] | Ivone Giffard-Mena[4] | Felipe Zapata[5] | Adam Roddy[6,7]

[1]Department of Biology, California State University Dominguez Hills, Carson, California 90747, USA

[2]University of California Institute for Mexico and the United States, University of California, Riverside, California 92521, USA

[3]Pinecrest Research Corporation, Oakland, California 94609, USA

[4]Facultad de Ciencias Marinas, Universidad Autónoma de Baja California, 22860 Ensenada, Baja California, México

[5]Department of Ecology and Evolutionary Biology and Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, California 90095, USA

[6]Institute of Environment, Department of Biological Sciences, Florida International University, Miami, Florida 33133, USA

[7]Department of Environmental Studies, New York University, New York, New York 10003, USA

Correspondence

Sonal Singhal, Department of Biology, California State University Dominguez Hills, Carson, California 90747, USA.
Email: sonal.singhal1@gmail.com

Abstract

Premise: Clines—or the geographic sorting of phenotypes across continual space—provide an opportunity to understand the interaction of dispersal, selection, and history in structuring polymorphisms.

Methods: In this study, we combine field-sampling, genetics, climatic analyses, and machine learning to understand a flower color polymorphism in the wide-ranging desert annual *Encelia farinosa*.

Results: We find evidence for replicated transitions in disk floret color from brown to yellow across spatial scales, with the most prominent cline stretching ~100 km from southwestern United States into México. Because population structure across the cline is minimal, selection is more likely than drift to have an important role in determining cline width.

Conclusions: Given that the cline aligns with a climatic transition but there is no evidence for pollinator preference for flower color, we hypothesize that floret color likely varies as a function of climatic conditions.

KEYWORDS

Asteraceae, bioclimatic analyses, clines, double-digest restriction-site associated DNA (ddRAD), North America

Many species exhibit clines, or continual change in phenotypes or genotypes across geographic space (Haldane, 1948; Endler, 1977). One way to characterize clines is by inferring cline width, or the geographic extent over which one form transitions into another. Some clines have incredibly narrow widths and can occur over just a few hundred meters, such as transitions between blue- and white-flowered varieties across ravines in the desert plant *Linanthus parryae* (A.Gray) Greene [Polemoniaceae] (Epling and Dobzhansky, 1942; Schemske and Bierzychudek, 2007). Others are continental in scale and can stretch over hundreds or thousands of kilometers, like many of the genetic and phenotypic transitions seen in modern humans (Novembre and Di Rienzo, 2009). What defines the width of a cline depends primarily on three factors, i.e., dispersal, selection, and history (Slatkin, 1973; Endler, 1977; Barton and Hewitt, 1985). First, greater dispersal across the cline leads to greater widths (Slatkin, 1973). Especially when clines fall in areas of low environmental suitability, dispersal through the cline is often lower than dispersal in the core of the range, further narrowing cline widths. Second, stronger selection against individuals dispersing across the cline decreases cline widths (Slatkin, 1973), whether this selection is mediated by the environment (Endler, 1977; DiVittorio et al., 2020) or independent of the environment, as can occur after hybridization (Barton and Gale, 1993). Finally, if clines are not yet at equilibrium—which can happen both

under secondary contact and during range expansions—time since the origin of the cline also influences width (Barton and Gale, 1993). Thus, the width of a cline can reveal the demography, adaptive significance, and evolutionary history of variation.

One such cline occurs in the desert brittlebush shrub (*Encelia farinosa* A.Gray ex Torr., Asteraceae). This wide-ranging species is common throughout its geographic distribution, which extends across the Mojave, Sonoran, and peninsular deserts in the southwestern United States and northwestern México. Although genetic structuring in the species is relatively modest (Fehlberg and Ranker, 2009; Fehlberg and Fehlberg, 2017), morphology and physiology vary throughout the range of the species (Sandquist and Ehleringer, 1996, 2003; Ehleringer and Sandquist, 2018). Two varieties have been characterized based on floret color: *Encelia farinosa* var. *farinosa* has yellow ray florets and yellow disk florets while *E. f.* var. *phenicodonta* (S.F.Blake) I.M.Johnst. has yellow ray florets and brown disk florets. Color appears to be genetically encoded and the patterns of inheritance follow a simple single-gene model, with brown being dominant to yellow disks (Kyhos, 1971).

These two varieties occur in sympatry in a few transitional regions (Kyhos, 1971), but throughout most of the species' range, one variety or the other dominates. While *E. f.* var. *phenicodonta* is primarily found in the Baja California peninsula, it also reaches north through the Colorado River basin and along the eastern coast of the Gulf of California. This distribution results in multiple transitional zones between the two varieties, which vary in estimated length from hundreds of kilometers along the Baja California peninsula to just a few kilometers over elevational gradients along the Colorado River (Kyhos, 1971). One possible explanation for this distribution is that floret color is adaptive, and the geographic segregation of these two varieties might reflect local adaptation to changing environmental conditions along the range (Kyhos, 1971; Sandquist and Ehleringer, 1996). In other plant species, flower color can affect fitness through both pollinator and non-pollinator effects (Rausher, 2008; Sobel and Streisfeld, 2013; Koski and Galloway, 2020; Sullivan and Koski, 2021). *Encelia farinosa* is pollinated by generalists—i.e., beetles, butterflies, and solitary bees (Kyhos, 1971; Clark, 1998)—which do not appear to discriminate between the two color morphs (Kyhos, 1971). Thus, how floret color might affect fitness in *E. farinosa* is unclear (Sandquist and Ehleringer, 1996).

In this study, we explore the roles of dispersal and selection in structuring flower color polymorphism in *E. farinosa*. To do so, we collected flower color data and genetic samples along a 1500-km transect through the major transition in disk floret color from southern California through the Baja California peninsula (Figure 1). Then, we analyzed genome-scale data for 46 individuals to determine if color polymorphism correlates with population structure. Next, we estimate the width and location of the genetic, phenotypic, and climate transitions to understand if and how these transitions correlate with each other. Finally, we place our results in a broader context by determining the range-wide distribution of the two varieties through machine learning of community-sourced images and by inferring the patterns of floret color evolution across the *Encelia* genus.

MATERIALS AND METHODS

Sampling

Encelia farinosa is a wide-ranging species, extending throughout the southwestern United States and northwestern México. Our sampling focused on the widest known transition in disk floret color (Kyhos, 1971), which occurs close to the border between the United States of America and México (Figure 1). Over the span of eight days in February 2016, we collected along a transect from Death Valley, California, United States to La Paz, Baja California Sur, México. Across 111 unique localities in this transect, we counted the number of plants in an approximately 200-m radius and recorded the proportion of plants that had brown disk florets. We concentrated our genetic sampling near the transition in color. We randomly sampled 28 localities for 1 to 3 individuals (total $n = 48$) into silica for genetic data collection.

Genetic data collection and processing

We used double-digest restriction-site associated DNA (ddRAD) to collect genetic data from *E. farinosa* individuals, as fully described by Singhal et al. (2021). Briefly, we first extracted DNA from silica-dried leaves using a DNeasy Plant Kit (catalog number 69104; QIAGEN, Germantown, Maryland, United States), digested the DNA using *PstI* and *MspI*, and size-selected DNA fragments of length 250 to 700 base pairs (bp) using a Pippin Prep (Sage Science, Beverly, Massachusetts, United States). We then prepared individually barcoded libraries following (Peterson et al., 2012). All libraries were pooled in equimolar amounts and then sequenced on a single 100-bp paired-end lane of Illumina HiSeq. 4000 at the Vincent Coates Genome Sequencing Laboratory (Berkeley, California, United States).

After demultiplexing the data, we first removed adapter sequence and regions with high sequencing error using trimmomatic version 36 (Bolger et al., 2014) and merged overlapping paired-end reads using PEAR version 0.9.8 (Zhang et al., 2014). We then created a pseudo-reference genome for our individuals. To do so, we first assembled each individual's reads into contigs using Velvet version 1.2.10 (Zerbino and Birney, 2008) and then identified homologous contigs across all individuals using VSEARCH version 2.9.1 (Rognes et al., 2016) with a 95% similarity threshold. The set of homologous contigs found across >50% individuals formed our pseudo-reference genome. Finally, we mapped cleaned reads to the pseudo-reference

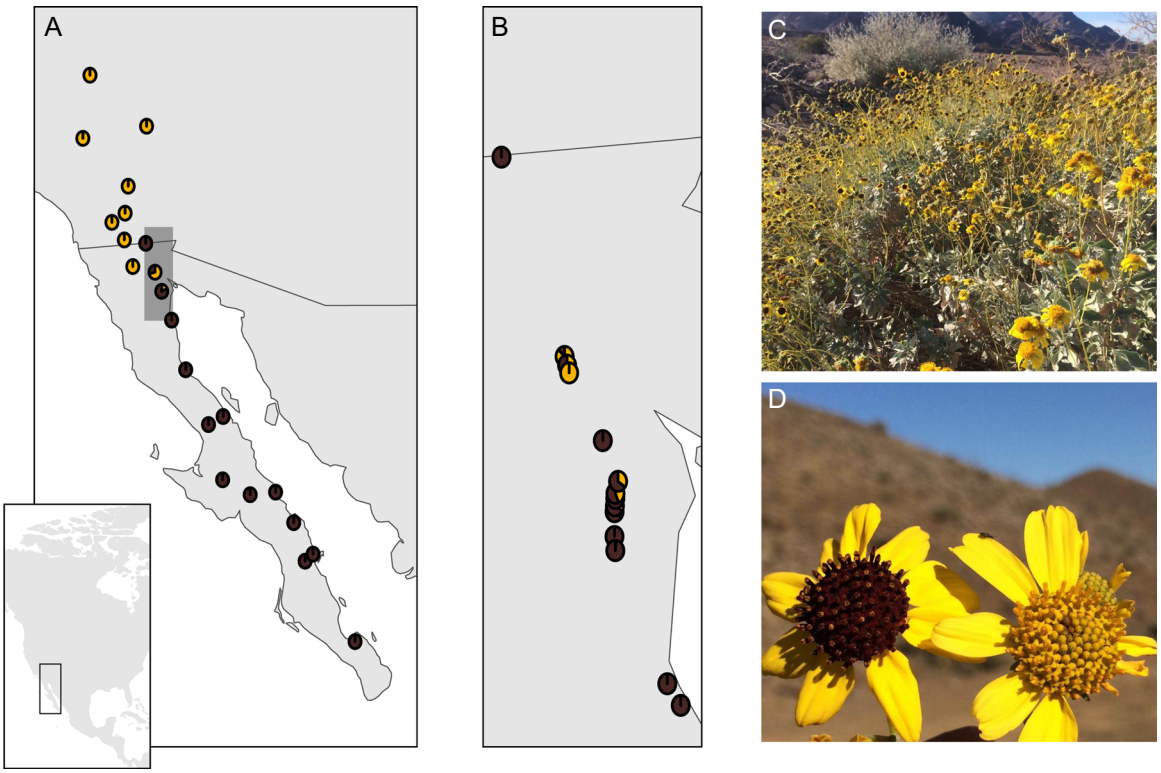

FIGURE 1 Variation in disk floret coloration of *Encelia farinosa* across the Southwestern United States to the Baja California peninsula in México. (A) Clinal transition from yellow to dark-brown disk floret. The shaded gray box marks the inset shown in (B). (B) A more detailed view of the transition at the cline center. (C) The yellow and dark-brown varieties of *E. farinosa* growing in sympatry, taken in Baja California on the coast of the Sea of Cortez near Puertocitos, México. (D) Floral heads from the two varieties of *E. farinosa*: Dark-brown and yellow. Photographs by C. DiVittorio.

genome using bwa version 0.7.17 (Li, 2013) and called variant and invariant sites using samtools version 1.5 (Li et al., 2009). We retained only those sites with a quality score >20 and for which missingness <50%. We additionally removed sites at which coverage was >5× the median coverage; such sites are likely to result from collapsed paralogs or other mis-assemblies. After filtering, two individuals were dropped due to high levels of missing data; our final genetic dataset consisted of 46 individuals (Appendix S1: Table S1).

Genetic analysis

Our genetic analyses focused on determining whether there was evidence of population structuring across the transition in disk floret color. First, we evaluated support for isolation-by-distance by comparing matrices of geographic and genetic distances between individuals using a Mantel test as implemented in vegan version 2.6 (Dixon, 2003). We calculated genetic distance between all pairwise combinations of individuals as inverse F_{ST} ($F_{ST}/[1 - F_{ST}]$ (Reich et al., 2009). For the next two analyses, we filtered our genotype set to retain only one site per ddRAD contig and only those sites with minor allele count ≥2 (Linck and Battey, 2019). Using this filtered variant set, we summarized patterns of genetic variation using a genetic principal component analysis (PCA) as implemented in adegenet version 2.1.10 (Jombart, 2008). Then, we used the model-based genetic clustering approach implemented in ADMIXTURE version 1.3.0 (Alexander et al., 2009) to determine the best model of population structure, running the program across one to five genetic clusters ($K = 1–5$). We used cross-validation to determine the best-fitting K value. Finally, using the full variant dataset, we used vcftools version 0.1.16 (Danecek et al., 2011) to calculate F_{ST} between clusters ($K = 2$) defined by an ADMIXTURE analysis and between color morphs.

Clinal analysis

We used HZAR version 0.2-5 (Derryberry et al., 2014) to characterize the clinal transitions in disk floret color, population ancestry, and climate across *E. farinose*. For population ancestry, we used the proportion estimated from ADMIXTURE under a model of two populations ($K = 2$) for the 46 individuals sampled for genetic analyses. For climate, we extracted values at our 111 unique localities across 19 bioclimatic variables at 30-seconds resolution available in WorldClim (Fick and Hijmans, 2017). We then summarized these values using a bioclimatic PCA as implemented in pcaMethods version 1.92.0 in R (Stacklies et al., 2007). We retained PC axes 1 and 2, which explained 34.5% and 28.1% of the variation, respectively. PC1 was dominated largely by annual mean temperature and annual mean

precipitation, while PC2 was dominated by seasonality and isothermality (Appendix S1: Table S2). Then, because our transect was not sampled at even intervals, we grouped localities by their nearest-neighbor distances (<5 km) to form 37 geographic groupings. For each of these geographic groups, we calculated average values for percent brown disk floret color, population ancestry, and climate variables.

Using HZAR, we then fitted a sigmoidal cline model to the flower, genetic, and climate data: $frequency = 1/(1 + e^{\frac{-4(x-center)}{width}})$ (Derryberry et al., 2014). This model defines the change in frequency across space as a function of both width—or, the geographic extent of the transition—and center—or, where in space the transition is located. We did not attempt to fit more parameter-rich models because of insufficient sampling in the cline tails. Then, for the 19 bioclimatic variables and the first two PC axes, we determined which variables most strongly correlated with the transition in flower color. To do so, we constructed a cline model where flower color varies in response to each climatic variable and measured model fit as the residual sum of squares between observed and predicted variables. Two climatic variables had much better model fit than the other variables (precipitation in driest quarter and in driest month; Appendix S1: Table S3), and they were highly correlated with each other ($\rho = 0.92$).

Finally, we tested if the clines had overlapping transitions (coincidence) by comparing model fits (as measured by Akaike Information Criterion [AIC]) for a model in which centers were allowed to vary across clines versus one in which centers were shared across clines. We tested for coincidence between color and climate clines and genetics and color clines. If clines are coincident, this suggests they are responding similarly to the same forces.

Mapping floret color variation across the broader *Encelia farinosa* range

Previous work on *E. farinosa* identified multiple other transitions between the brown and yellow morphs (Kyhos, 1971; Sandquist and Ehleringer, 1996; Fehlberg and Fehlberg, 2017). To map these transitions, we relied on community-generated images available on iNaturalist (https://www.inaturalist.org/). On 19 December 2023, we downloaded all research-grade *Encelia farinosa* image observations (n = 17,301). We developed a machine-learning model to classify disk floret color across this large number of images, inspired by the approach of Hantak et al. (2022).

First, we trained and validated the model using a smaller set of manually annotated images. Two independent researchers (IG and SS) categorized images (n = 1856) as either having no florets, brown disk florets, or yellow disk florets using the program ImageAnt (Hantak et al., 2022). Concordance between the two researchers was 99.3%; we used these concordant images to train the models. Images were resized, randomly flipped, and color normalized before being loaded. Because we had fewer images of brown florets relative to yellow florets, we additionally applied an oversampling procedure. Then, we used transfer learning to train a convolutional neural network, as implemented in PyTorch Lightning. Here, we applied the efficientnet-b4 model (Tan and Le, 2019). We ran a four-fold cross-validation across four learning rates (0.1, 0.01, 0.001, or 0.0001). Each training loop was run across 100 epochs. We then identified the most accurate learning rate by applying these models to a set of 500 independent images again manually annotated with ImageAnt. Using the learning rate that generated the highest accuracy on this test set (learning rate = 0.01), we classified all remaining images using the average prediction across the four-fold models. We summarized our results by binning the predicted data into a spatial grid using the sf version 1.0-14 package in R (Pebesma, 2018).

Mapping floret color variation across the genus

We additionally investigated patterns of floret color across *Encelia*. By manually annotating iNaturalist images (approximately 10–20 photos per species) and summarizing field guides (Rebman and Roberts, 2012), we identified the disk floret color of each species. Only *E. farinosa* and *E. californica* Nutt. are known to be polymorphic. Using the species tree from (Singhal et al., 2021), we reconstructed disk floret color across the genus using 100 simulations of the stochastic character mapping algorithm implemented in phytools version 2.0 (Revell, 2024). Finally, we mapped the geographic distributions of the yellow- vs. brown-colored florets using research-grade observations downloaded from Global Biodiversity Information Facility (GBIF; https://www.gbif.org/), as described by Singhal et al. (2021).

Data analysis and visualization

We used R version 4.3.1—including the ggplot version 3.4.4 and the tidyverse version 2.0.0 packages—for data analysis and visualization (Wickham, 2016; Wickham et al., 2019; R Core Team, 2021).

RESULTS

Population structure

After assembly of our genetic data and filtering for quality and missingness, we retained 116K variant sites across 81K loci in 46 individuals. We then used these data to test for population structure between the yellow and brown morphs, finding only weak evidence for structure. Across the species, we see significant evidence of isolation-by-distance (IBD; $r = 0.30$; p-value = 0.001; Figure 2B). The IBD slopes for within-morph comparisons overlap with the IBD slopes for between-morph

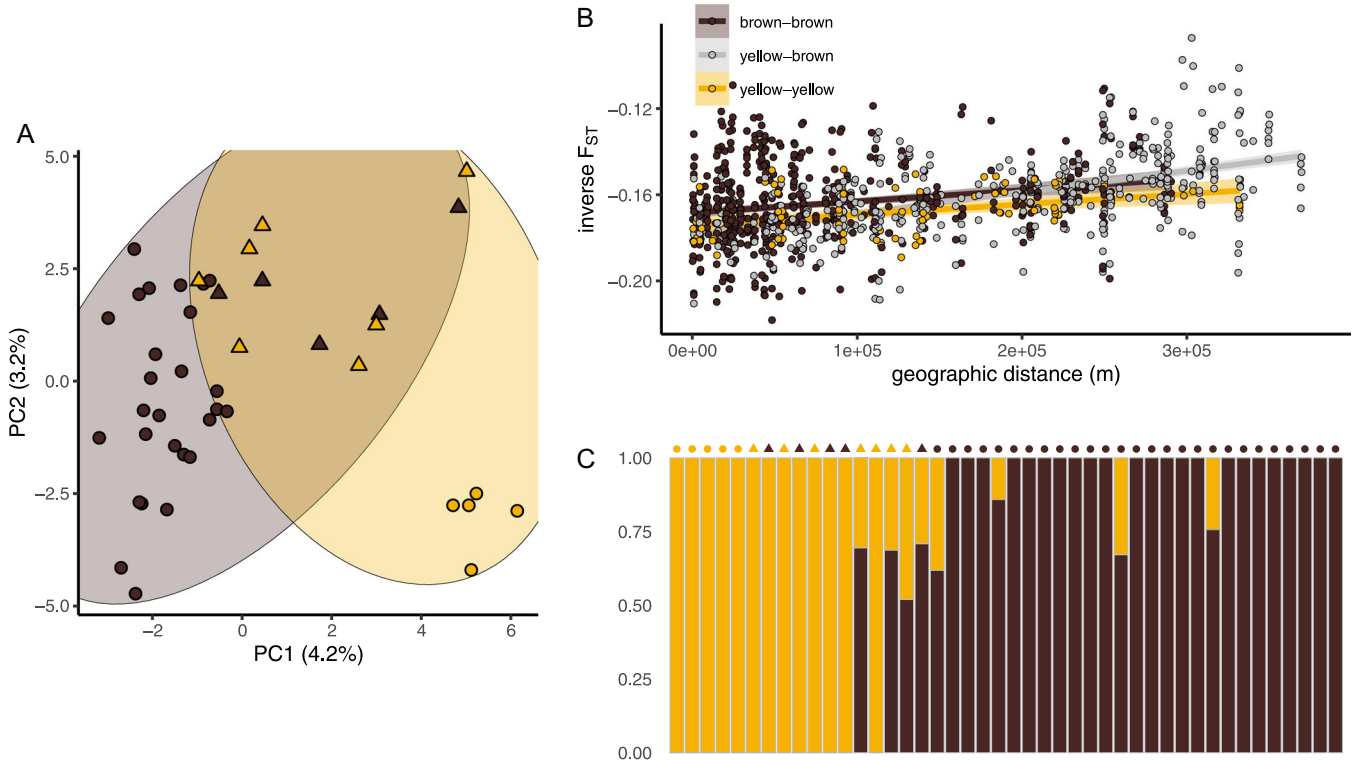

FIGURE 2 Patterns of population structure across the yellow-brown color transition in *Encelia farinosa* (n = 46 individuals). (A) Principal component analysis of genomic variation shows that individuals with brown disk florets (shown in brown) are slightly differentiated from those with yellow disk florets (shown in yellow). Individuals found in the middle of the transition (see Figure 1B) are shown as triangles. (B) Pairwise geographic distance and genetic distance (as measured by inverse F_{ST}) across all combinations of genetically sampled individuals. The species shows evidence for isolation-by-distance (IBD), and IBD patterns are similar across disk floret color comparisons. (C) Population clustering results for $K = 2$ results in two genetic groups that somewhat map to disk floret color (shown above the cluster plot). Individuals are ordered from north to south, and individuals in the transition are represented by triangles.

comparisons, suggesting the two morphs are not separated by a sharp genetic discontinuity. Further, while the brown and yellow morphs show evidence for segregation along the two axes of the genetic PCA, these two axes explained a relatively low proportion of the total variation (<5%; Figure 2A). Finally, genetic clustering results from ADMIXTURE suggested that a single-cluster model ($K = 1$) best fit the data (Appendix S1: Table S4). Results from a two-cluster model ($K = 2$), however, show that the individuals are assigned to two populations that roughly align to the morph identity (Figure 2C). The Weir-Cockerham weighted F_{ST} between both the two genetic clusters is low ($F_{ST} = 0.018$), and F_{ST} between color morphs is even lower ($F_{ST} = 0.006$).

Clinal analysis

Cline width and center varied across the transitions in color, genetics, and climate in our north-south transect (Figure 3). Climate PC1—which mainly reflected annual mean temperature and precipitation—showed no clear pattern across the transect (Appendix S2: Figure S1), so our results here focus on climate PC2, which mainly reflected seasonality and isothermality. Brown florets are associated with reduced seasonality, greater isothermality, warmer temperatures during cold periods, and lower precipitation during dry periods (Appendix S2: Figure S2). The environmental metric that best predicts patterns of flower color is precipitation in the driest quarter (Appendix S1: Table S3; Figure 3B).

The cline width for disk floret color was estimated to be 132 km (95% credibility range: 123 to 143 km), slightly wider than the genetic cline ($w = 100$ km; 95% credibility range: 52 to 198 km) and much narrower than the climate transition ($w = 998$ km for PC2; 95% credibility range: 664 to 1599 km). We found no evidence for coincidence of clines in color and genetics ($\Delta AIC = 19.51$; the constrained model has a <0.1% probability of being a better model than the unconstrained model) but some evidence for coincidence of clines in color and climate ($\Delta AIC = 1.97$; the constrained model has a 63% probability of being a better model than the unconstrained model).

Floret color variation across the ranges of *Encelia farinosa* and the genus *Encelia*

Our machine-learning model had an accuracy of 98.8% when including all three possible predictions (no flowers, brown disk florets, or yellow disk florets) and 96.6% when

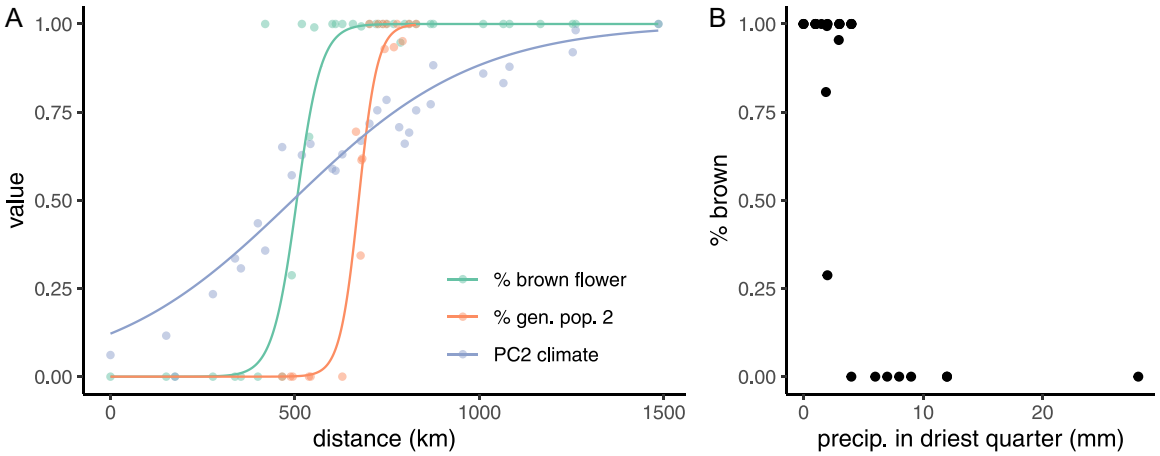

FIGURE 3 (A) Estimates for clines along a north-south transect in *Encelia farinosa*. Shown is percentage of flowers with brown disk florets (% brown flower; green), ancestry proportion for genetic population 2 (% gen. pop. 2; orange), and normalized value for principal component 2 for climate (PC2 climate; blue). Points represent empirical values. The flower color cline is not coincident with the genetic cline, and we see support for coincidence with the transition in climate. (B) Precipitation in the driest quarter best predicts the percentage of brown flowers at a given locality.

considering disk floret color only. Overall, 44.5% of our images were predicted to contain flowers, of which 8.3% were predicted to have brown disk florets. As has been suggested by previous studies (Kyhos, 1971; Sandquist and Ehleringer, 1996; Fehlberg and Fehlberg, 2017), the brown morph was found primarily in southern Baja California peninsula, along the western coast of Sonora, México, and along the rivers of the western United States (Figure 4). Additional small isolated pockets of the brown morph occurred throughout the range of *E. farinosa*.

Mapping of disk floret color across the *Encelia* phylogeny showed that the ancestral state was as likely to be yellow as brown and that there were three transitions in flower color across the clade (Figure 5A). *Encelia farinosa* is the only species found in both the northern and southern distributions of *Encelia*, and the primary yellow-brown transition seen in *E. farinosa* loosely maps to the distribution of these floret colors genus-wide (Figure 5B).

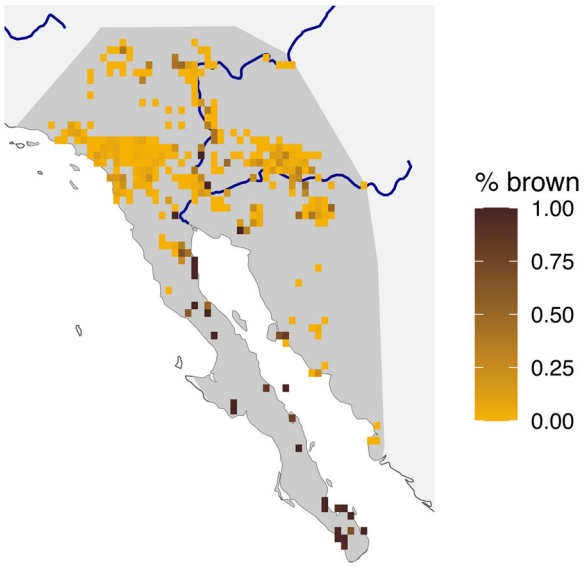

FIGURE 4 Distribution of the brown and yellow morphs of *Encelia farinosa* across its entire range (shown in gray). Images from iNaturalist (n = 17,301) were classified as showing brown or yellow disk florets using a convolutional neural network machine; observations were then spatially binned and summarized. Waterways shown in blue outline. While the major transition between the two morphs occurs on the Baja California peninsula, the brown morph is found throughout the *E. farinosa* range.

DISCUSSION

Within the wide-ranging desert plant species *Encelia farinosa*, we describe a 100-kilometer cline in disk floret color from yellow in the north to brown in the south (Figure 1). Below, we discuss what the genetic and trait data we collected across this transition can tell us about how dispersal and selection might be structuring the origin and maintenance of this polymorphism.

Genetic differentiation in *Encelia farinosa*

No matter how a cline formed, the extent of genetic differentiation across a cline can reveal the relative balance of gene flow and genetic drift in shaping the polymorphism (Slatkin, 1973; Endler, 1977; Vasemägi, 2006). If the polymorphism has arisen largely due to neutral sorting by drift or via secondary contact of previously isolated forms, we would expect that population structure would align with the spatial patterns of the polymorphism (Streisfeld and Kohn, 2005; Schemske and Bierzychudek, 2007). Our continuous sampling through the widest transition in flower color revealed that population structure was weak. While sampled individuals largely fell out by color in genotypic

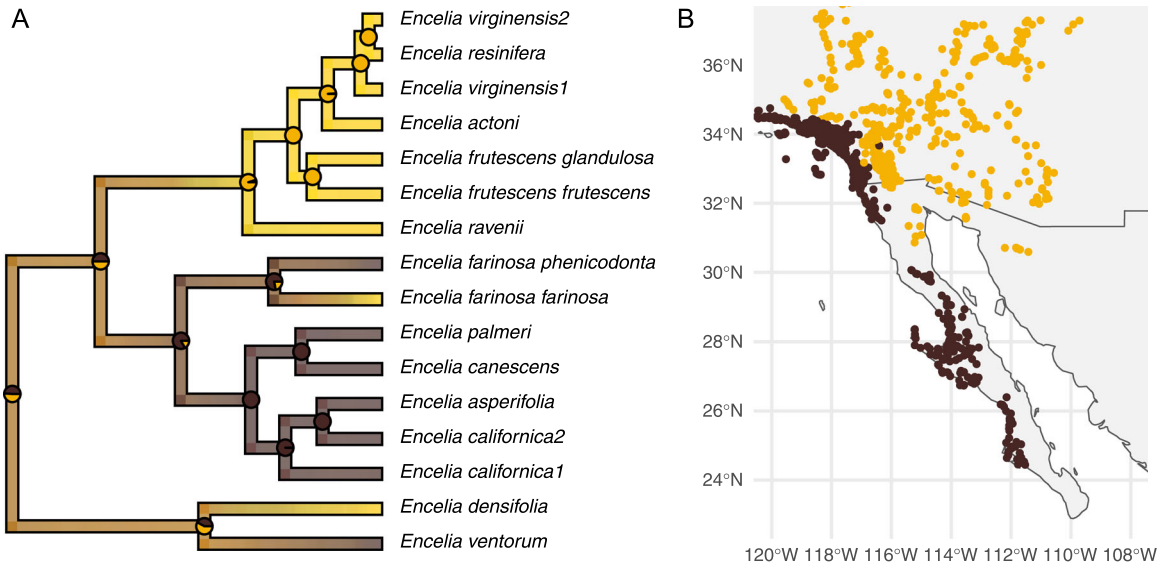

FIGURE 5 Disk floret coloration across the *Encelia* radiation. (A) The species tree topology of *Encelia*, with disk floret color reconstruction mapped across nodes. The ancestral color is as likely to be brown as yellow. (B) Geographic distribution of yellow-colored versus brown-colored species across all *Encelia* species except *E. farinosa* (not shown for clarity) and *E. canescens* Lam. (not shown because it is found in South America). *Encelia farinosa* is the only species found in both the northern and southern distributions of *Encelia*, and the primary yellow-brown transition in *E. farinosa* loosely maps to the distribution of these colors genus-wide.

space (Figure 2A), there is only a very slight discontinuity in isolation-by-distance patterns across the color transition (Figure 2B). Further, genetic clustering approaches suggest a model of just one genetic cluster that best fits the data (Appendix S1: Table S4), and genetic differentiation as estimated by F_{ST} is extremely low between the two flower morphs ($F_{ST} = 0.006$). Thus, if this cline formed due to secondary contact, then either extensive gene flow eroded genetic divergence between the forms or the forms were only weakly differentiated to start. Similarly, while drift may be acting to narrow the cline (Polechová and Barton, 2011), the low levels of population structure suggest drift is not a primary force structuring this polymorphism.

Low levels of population structure across the cline are not surprising given *E. farinosa* biology. Like most of the species in the genus, *Encelia farinosa* is an obligate outcrosser and is pollinated by generalists (Clark, 1998), with pollinators exhibiting little discrimination between the two color morphs (Kyhos, 1971). Further, *E. farinosa* is distributed continuously throughout the transition between the morphs. Thus, gene flow across the transition and through the range is likely high, which would slow the formation of population structure. Indeed, previous work on *Encelia farinosa* found that, despite high levels of genetic variability, most of the variation is shared across the geographic range of the species (Fehlberg and Ranker, 2009; Fehlberg and Fehlberg, 2017).

This low-level population structure also follows from the likely genetic architecture of disk floret color. Previous crossing experiments suggest that floret color is inherited due to variation at one locus, with the brown color having a dominant pattern of inheritance (Kyhos, 1971). Even if strong selection was maintaining the color polymorphism, selection would act solely on one locus and its linkage block (Slatkin, 1975). Especially in the absence of assortative mating across morphs, this selection would be unlikely to precipitate genome-wide differentiation (Schluter, 2000; Kirkpatrick and Ravigné, 2002).

Natural selection and the floret color polymorphism

Natural selection likely has an important role in maintaining the floret color cline in *Encelia farinosa*, especially given the cline's replicated occurrence across scales (Endler, 1977; Antoniazza et al., 2010). We consider the potential selective benefit of disk floret color by exploring cline width and phenotype-environment associations.

If the rate of dispersal is known, then selection can be inferred based on the width of the cline. Though we do not know dispersal rates across this cline, studies of other *Encelia* species have shown that pollen can disperse at least 200 m (DiVittorio et al., 2020). Under the gradient model in which the dispersal distance is much smaller than the cline width (Endler, 1977), the relationship between selection (b), cline width (w), and dispersal (l) is: $b = l^2 \times \left(\frac{2.4}{w}\right)^3$. With an upper dispersal distance of 200 m, b is approximately 2e-7. If mean dispersal is lower than 200 m, then b would be even lower. However, coarse spatial data from (Kyhos, 1971) suggest some of the clines in floret color might be only one to two kilometers wide, such as transitions along elevational gradients along the Colorado River. Again assuming

dispersal at ~200 m, and with a two kilometer cline width, selection on floret color would be on the order of ~0.07 in this part of the *E. farinosa* range. Thus, it appears the selection gradient on floret color is highly contingent on the associated environmental gradient.

There are three possible sources of selection in *E. farinosa*: (1) pollinator preference; (2) genetic linkage of flower color to another selected trait; and (3) the abiotic environment. First, in many plant species (Rausher, 2008; Caruso et al., 2019), pollinators exhibit preference for different floral morphs, thus helping maintain color polymorphisms. Both the brown and yellow morphs of *E. farinosa* absorb long-wave ultraviolet radiation (Clark and Sanders, 1986), and the generalist pollinators of *E. farinosa* seem to have equal preference for both disks (Kyhos, 1971). Pollinator preference is thus unlikely to be structuring this polymorphism. Second, the loci underpinning floral color might be physically linked to loci underpinning another trait under direct selection (Rausher, 2008). Flower color variation would thus emerge as a correlated response to selection on the other trait. For example, morphological data suggest that other traits might co-vary with flower color, such as trichome density and leaf shape (Singhal et al., 2021).

Third, this polymorphism could be structured by the abiotic environment, whether through direct or pleiotropic effects. Floret color could affect physiological performance directly (Harder and Barrett, 2006; Sapir et al., 2021). For example, darker flowers can have higher fitness than lighter flowers because they better protect against UV damage (Thompson et al., 1972; Koski and Ashman, 2015). Alternatively, flower color could vary as a pleiotropic indicator of systemic physiological responses to abiotic stress (Harder and Barrett, 2006). Based on patterns in other species (Tanaka et al., 2008; Sobel and Streisfeld, 2013), the purple-brown color in *E. farinosa* likely results from increased production of the flavonoid anthocyanin (Grotewold, 2006; Wessinger and Rausher, 2012). Anthocyanins and other flavonoids are typically upregulated in response to stress (Clegg and Durbin, 2000; Harder and Barrett, 2006; Rausher, 2008; Wessinger and Rausher, 2012); plant organs with higher flavonoid concentrations are more freeze, drought, and heat tolerant (Chalker-Scott, 1999; Winkel-Shirley, 2002; Wessinger and Rausher, 2012). Thus, flower color sometimes varies along gradients in abiotic stressors (Dick et al., 2011; Berardi et al., 2016), whether it is the direct target of selection or the result of pleiotropy.

The patterns of floret color in *E. farinosa* align somewhat with predictions based on environmental gradients. The transition in disk floret color is coincident with the transition in climate as summarized by PC2 (Figure 3). In particular, the brown morph is more common in areas that experience drier dry seasons (Figure 3B), suggesting that anthocyanins might be upregulated in response to drought. However, the yellow morph is more common in areas that experience freezing (Appendix S2: Figure S2), which, while counter to physiological predictions, matches experimental trials that show the yellow morph tolerates freezing temperatures better than the brown morph (Sandquist and Ehleringer, 1996). Notably, this association between floret color and climate is shared among the genus *Encelia* more broadly. All the species found in the colder, northern part of the *Encelia* distribution have yellow disk florets, whereas those found in the hotter, drier, southern part of the distribution have brown disk florets (Figure 5). In *Encelia farinosa*, which is the most geographically widespread species in the genus, the spatial extent of the yellow-brown polymorphism maps nearly perfectly to the geographic variation in disk color genus-wide.

In addition to this broad spatial segregation of flower color across the entire genus and within *E. farinosa*, isolated pockets of the brown morph are found within the range of the yellow morph in *E. farinosa* (Figure 4). These isolated occurrences occur across narrow geographic scales (anywhere from a few to tens of kilometers), and thus they cannot be explained by the broad-scale clines in drought and freezing temperatures seen across the primary color morph transition seen in the whole species *E. farinosa* (Figure 3). Kyhos (1971) argued that the one commonality among these restricted occurrences of the brown morph is the presence of gradients in water deficit, with the brown morph being found in what are thought to be wetter, presumably less stressful habitats. This hypothesis runs counter to the idea that the brown morph is prevalent due to the upregulation of anthocyanins in areas of stress, suggesting that either water deficits are not stressful for *E. farinosa* or that anthocyanin production is being driven by other abiotic conditions.

The demographic history of *E. farinosa* might also explain the occurrence of brown morphs in the range of the yellow morph. *Encelia farinosa* is thought to have undergone a rapid population expansion during the post-glacial spread of deserts (Fehlberg and Ranker, 2009; Fehlberg and Fehlberg, 2017). Genetic data suggest that historically stable populations were found along the Colorado River, the Sonoran Coast, and some parts of Baja California (Fehlberg and Ranker, 2009; Fehlberg and Fehlberg, 2017). This geographic distribution directly mirrors that of the modern distribution of the brown morph. Given that the ancestral disk floret in *E. farinosa* was likely brown (Figure 5), these isolated populations of brown morphs in the midst of yellow floret populations might be relictual populations, which have managed to persist and thrive in present-day climatic conditions. Indeed, physiological data from a common garden suggest that the brown morph is fully capable of surviving most of the abiotic challenges of the northern range except for freezing temperatures (Sandquist and Ehleringer, 1996).

Origins of floret color variation in *Encelia*

There have been multiple transitions among yellow and brown disk florets in *Encelia* (Figure 5), A pattern seen for other traits in the genus (e.g., high trichome density; Singhal et al., 2021). Such repeated evolution could have three possible sources. First, each transition might be due to

independent de novo mutations to genes in the flavonoid synthesis pathway (Wessinger and Rausher, 2012; Sobel and Streisfeld, 2013). Second, perhaps the ancestral floret color condition was polymorphic, and the rapid radiation of *Encelia* led to differential sorting across species (Lee and Coop, 2019). Third, given the propensity of *Encelia* to hybridize (Clark, 1998; Singhal et al., 2021), yellow disk florets may have arisen once and then spread to *Encelia farinosa* via selective introgression. Distinguishing between these three possibilities will require uncovering the genetic basis of disk floret coloration.

CONCLUSIONS

This study characterizes a polymorphism in disk floret coloration among the wide-ranging desert species *Encelia farinosa* (Figure 1), showing how the polymorphism is replicated across varying spatial scales ranging from thousands of kilometers to just a few kilometers (Figure 4) and across the genus as a whole (Figure 5). Although we do not yet know the exact mechanisms maintaining this polymorphism, our analyses suggest the polymorphism is coupled with climatic transitions (Figure 3). Given the replication across scales and species and the linkage to abiotic conditions, the disk floret polymorphism in *Encelia* is a powerful system to understand genetic and adaptive basis for the repeated evolution of phenotypes.

AUTHOR CONTRIBUTIONS

S.S.: Conceptualization, analysis, investigation, resources, data curation, writing—original draft, writing—review & editing, visualization, and funding acquisition. C.D.V.: Conceptualization, investigation, resources, writing—review & editing, and funding acquisition. C.J.: Investigation, analysis, data curation, and visualization. I.I.: Investigation. A.W.: Investigation and analysis. I.G.M.: Resources, writing—review & editing, and funding acquisition. F.Z.: Conceptualization and writing—review & editing. A.R.: Conceptualization, Wwiting—review & editing and funding acquisition.

ACKNOWLEDGMENTS

We gratefully acknowledge funding from UC MEXUS to C.T.D., I.G.M., S.S.; a CSUDH RSCA grant to S.S.; LS-AMP Student Support to I.I. and C.J.; and a National Science Foundation award DEB 2023979 to C.J. and A.W. For logistical and technical support, we thank Erica Bree Rosenblum and Lydia Smith. We also thank the Associate Editor and two anonymous reviewers for their constructive comments on earlier versions of this manuscript.

DATA AVAILABILITY STATEMENT

Genomic data are available at NCBI BioProject PRJNA 1109372 (https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1109372). Scripts used for data analysis and visualization and data are available at GitHub: https://github.com/singhal/farinosa.

ORCID

Sonal Singhal http://orcid.org/0000-0001-5407-5567
Felipe Zapata http://orcid.org/0000-0002-9386-0573

REFERENCES

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664.

Antoniazza, S., R. Burri, L. Fumagalli, J. Goudet, and A. Roulin. 2010. Local adaptation maintains clinal variation in melanin-based coloration of European barn owls (*Tyto alba*). *Evolution* 64: 1944–1954.

Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. *In* R. G. Harrison [ed.], Hybrid zones and the evolutionary process, 13–45. Oxford Academic, New York, New York, USA.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* 16: 113–148.

Berardi, A. E., P. D. Fields, J. L. Abbate, and D. R. Taylor. 2016. Elevational divergence and clinal variation in floral color and leaf chemistry in *Silene vulgaris*. *American Journal of Botany* 103: 1508–1523.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Caruso, C. M., K. E. Eisen, R. A. Martin, and N. Sletvold. 2019. A meta-analysis of the agents of selection on floral traits. *Evolution* 73: 4–14.

Chalker-Scott, L. 1999. Environmental significance of anthocyanins in plant stress responses. *Photochemistry and Photobiology* 70: 1–9.

Clark, C. 1998. Phylogeny and adaptation in the *Encelia* alliance (Asteraceae: Heliantheae). *Aliso* 17: 89–98.

Clark, C., and D. L. Sanders. 1986. Floral ultraviolet in the *Encelia* alliance (Asteraceae: Heliantheae). *Madroño* 33: 130–135.

Clegg, M. T., and M. L. Durbin. 2000. Flower color variation: A model for the experimental study of evolution. *Proceedings of the National Academy of Sciences. USA* 97: 7016–7023.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

Derryberry, E. P., G. E. Derryberry, J. M. Maley, and R. T. Brumfield. 2014. HZAR: Hybrid zone analysis using an R software package. *Molecular Ecology Resources* 14: 652–663.

Dick, C. A., J. Buenrostro, T. Butler, M. L. Carlson, D. J. Kliebenstein, and J. B. Whittall. 2011. Arctic mustard flower color polymorphism controlled by petal-specific downregulation at the threshold of the anthocyanin biosynthetic pathway. *PLoS One* 6: e18230.

DiVittorio, C. T., S. Singhal, A. B. Roddy, F. Zapata, D. D. Ackerly, B. G. Baldwin, C. R. Brodersen, et al. 2020. Natural selection maintains species despite frequent hybridization in the desert shrub *Encelia*. *Proceedings of the National Academy of Sciences, USA* 117: 33373–33383.

Dixon, P. 2003. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* 14: 927–930.

Ehleringer, J. R., and D. R. Sandquist. 2018. A tale of ENSO, PDO, and increasing aridity impacts on drought-deciduous shrubs in the Death Valley region. *Oecologia* 187: 879–895.

Endler, J. A. 1977. Geographic variation, speciation, and clines. Princeton University Press, Princeton, New Jersey, USA.

Epling, C., and T. Dobzhansky. 1942. Genetics of natural populations. VI. Microgeographic races in *Linanthus parryae*. *Genetics* 27: 317–332.

Fehlberg, S. D., and K. M. Fehlberg. 2017. Spatial genetic structure in brittlebush (*Encelia farinosa*, Asteraceae) in the southwestern deserts of North America: A comparison of nuclear and chloroplast DNA sequences. *Plant Systematics and Evolution* 303: 1367–1382.

Fehlberg, S. D., and T. A. Ranker. 2009. Evolutionary history and phylogeography of *Encelia farinosa* (Asteraceae) from the Sonoran, Mojave, and peninsular deserts. *Molecular Phylogenetics and Evolution* 50: 326–335.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37: 4302–4315.

Grotewold, E. 2006. The genetics and biochemistry of floral pigments. *Annual Review of Plant Biology* 57: 761–780.
Haldane, J. B. S. 1948. The theory of a cline. *Journal of Genetics* 48: 277–284.
Hantak, M. M., R. P. Guralnick, A. Zare, and B. J. Stucky. 2022. Computer vision for assessing species color pattern variation from web-based community science images. *iScience* 25: 104784.
Harder, L. D., and S. C. H. Barrett. 2006. Ecology and evolution of flowers. Oxford University Press, Oxford, UK.
Jombart, T. 2008. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405.
Kirkpatrick, M., and V. Ravigné. 2002. Speciation by natural and sexual selection: Models and experiments. *American Naturalist* 159 Suppl 3: S22–S35.
Koski, M. H., and T.-L. Ashman. 2015. Floral pigmentation patterns provide an example of Gloger's rule in plants. *Nature Plants* 1: 14007.
Koski, M. H., and L. F. Galloway. 2020. Geographic variation in floral color and reflectance correlates with temperature and colonization history. *Frontiers in Plant Science* 11: 991.
Kyhos, D. W. 1971. Evidence of different adaptations of flower color variants of *Encelia farinosa* (Compositae). *Madroño* 21: 49–61.
Lee, K. M., and G. Coop. 2019. Population genomics perspectives on convergent adaptation. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 374: 20180236.
Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303.3997. Website: https://doi.org/10.48550/arXiv.1303.3997
Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.
Linck, E., and C. J. Battey. 2019. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources* 19: 639–647.
Novembre, J., and A. Di Rienzo. 2009. Spatial patterns of variation due to natural selection in humans. *Nature Reviews Genetics* 10: 745–755.
Pebesma, E. 2018. Simple features for R: Standardized support for spatial vector data. *The R Journal* 10(1): 439–446.
Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7: e37135.
Polechová, J., and N. Barton. 2011. Genetic drift widens the expected cline but narrows the expected cline width. *Genetics* 189: 227–235.
R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website: https://www.R-project.org/
Rausher, M. D. 2008. Evolutionary transitions in floral color. *International Journal of Plant Sciences* 169: 7–21.
Rebman, J. P., and N. C. Roberts. 2012. Baja California plant field guide. Sunbelt Publications, Chula Vista, California, USA.
Reich, D., K. Thangaraj, N. Patterson, A. L. Price, and L. Singh. 2009. Reconstructing Indian population history. *Nature* 461: 489–494.
Revell, L. J. 2024. phytools 2.0: An updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* 12: e16505.
Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahé. 2016. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* 4: e2584.
Sandquist, D. R., and J. R. Ehleringer. 1996. Potential adaptability and constraints of response to changing climates for *Encelia farinosa* var. *phenicodonta* from southern Baja California, *Mexico*. *Madroño* 43: 465–478.
Sandquist, D. R., and J. R. Ehleringer. 2003. Carbon isotope discrimination differences within and between contrasting populations of *Encelia farinosa* raised under common-environment conditions. *Oecologia* 134: 463–470.
Sapir, Y., M. K. Gallagher, and E. Senden. 2021. What maintains flower colour variation within populations? *Trends in Ecology & Evolution* 36: 507–519.
Schemske, D. W., and P. Bierzychudek. 2007. Spatial differentiation for flower color in the desert annual *Linanthus parryae*: Was Wright right? *Evolution* 61: 2528–2543.
Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, Oxford, UK.
Singhal, S., A. B. Roddy, C. DiVittorio, A. Sanchez-Amaya, C. L. Henriquez, C. R. Brodersen, S. Fehlberg, and F. Zapata. 2021. Diversification, disparification and hybridization in the desert shrubs *Encelia*. *New Phytologist* 230: 1228–1241.
Slatkin, M. 1973. Gene flow and selection in a cline. *Genetics* 75: 733–756.
Slatkin, M. 1975. Gene flow and selection in a two-locus system. *Genetics* 81: 787–802.
Sobel, J. M., and M. A. Streisfeld. 2013. Flower color as a model system for studies of plant evo-devo. *Frontiers in Plant Science* 4: 321.
Stacklies, W., H. Redestig, M. Scholz, D. Walther, and J. Selbig. 2007. pcaMethods—a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* 23: 1164–1167.
Streisfeld, M. A., and J. R. Kohn. 2005. Contrasting patterns of floral and molecular variation across a cline in *Mimulus aurantiacus*. *Evolution* 59: 2548–2559.
Sullivan, C. N., and M. H. Koski. 2021. The effects of climate change on floral anthocyanin polymorphisms. *Proceedings of the Royal Society of London, B, Biological sciences* 288: 20202693.
Tan, M., and Q. Le. 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. *In* K. Chaudhuri and R. Salakhutdinov [eds.], Proceedings of the 36th International Conference on Machine Learning. *PMLR* 97: 6105–6114.
Tanaka, Y., N. Sasaki, and A. Ohmiya. 2008. Biosynthesis of plant pigments: Anthocyanins, betalains and carotenoids. *Plant Journal* 54: 733–749.
Thompson, W. R., J. Meinwald, D. Aneshansley, and T. Eisner. 1972. Flavonols: Pigments responsible for ultraviolet absorption in nectar guide of flower. *Science* 177: 528–530.
Vasemägi, A. 2006. The adaptive hypothesis of clinal variation revisited: Single-locus clines as a result of spatially restricted gene flow. *Genetics* 173: 2411–2414.
Wessinger, C. A., and M. D. Rausher. 2012. Lessons from flower colour evolution on targets of selection. *Journal of Experimental Botany* 63: 5741–5749.
Wickham, H. 2016. ggplot2: Elegant graphics for data analysis. Springer Verlag, Berlin, Germany.
Wickham, H., M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, et al. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4: 1686.
Winkel-Shirley, B. 2002. Biosynthesis of flavonoids and effects of stress. *Current Opinion in Plant Biology* 5: 218–223.
Zerbino, D. R., and E. Birney. 2008. Velvet: Algorithms for *de novo* short read assembly using *de Bruijn* graphs. *Genome Research* 18: 821–829.
Zhang, J., K. Kobert, T. Flouri, and A. Stamatakis. 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30: 614–620.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Supporting tables.

Appendix S2. Supporting figures.

How to cite this article: Singhal, S., C. DiVittorio, C. Jones, I. Ixta, A. Widmann, I. Giffard-Mena, F. Zapata, and A. Roddy. 2024. Population structure and natural selection across a flower color polymorphism in the desert plant *Encelia farinosa*. *American Journal of Botany* 111(10): e16413. https://doi.org/10.1002/ajb2.16413