Title

Evolution of a Specialized Herbivore Community: Biogeography, Speciation, and Balancing Selection in North American Deserts

Author

O'Connor, Timothy Kevin

Publication Date

2019

Peer reviewed|Thesis/dissertation

Evolution of a Specialized Herbivore Community: Biogeography, Speciation, and Balancing Selection in North American Deserts

by

Timothy K. O'Connor

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Noah Whiteman, Chair
Professor Erica Bree Rosenblum
Professor Rauri Bowie
Professor Michael Nachman

Fall 2019

Evolution of a Specialized Herbivore Community: Biogeography, Speciation, and Balancing Selection in North American Deserts

by Timothy K. O'Connor

Abstract

Evolution of a Specialized Herbivore Community: Biogeography, Speciation, and Balancing Selection in North American Deserts

by

Timothy K. O'Connor

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Professor Noah Whiteman, Chair

The many facets of diversity are an enduring source of fascination for biologists. In my dissertation, I present three case studies exploring how species interactions affect the generation, distribution, and maintenance of biological diversity.

My work centers on a guild of specialized herbivores that consume creosote bush (*Larrea tridentata*) in the warm deserts of North America. The nominal species *L. tridentata* comprises three insipient species that arose through autopolyploidy (whole genome duplication without hybridization). Speciation by polyploidy can have manifold consequences for plants. One common outcome is that cytotypes (plants of different ploidy level) geographically segregate. Indeed, the diploid, tetraploid, and hexaploid cytotypes of creosote bush are parapatrically distributed in the Chihuahuan, Sonoran, and Mojave Deserts, respectively. A second outcome is that the phenotypic changes accompanying polyploidy alter interactions with other species, including mutualists, competitors, and consumers. My first two chapters consider how the consequences of polyploidy affect the biogeography and genetic divergence of specialized herbivores.

In Chapter 1, I surveyed the diversity of gall midges in the genus *Asphondylia* across the full distribution of creosote bush (>2,300 km extent). I found that many gall species (6/17, including two new species) are specialized on either diploid or polyploid host plants. These species preferentially attack just one host cytotype even where diploid and tetraploid creosote bush commingle. The contact zone between creosote bush cytotypes is thus a biotically-mediated dispersal barrier, irrespective of abiotic factors or physical distance. Because more gall midges species specialize on polyploid host plants than on diploids, the distribution of creosote bush cytotypes structures a gradient in gall midge richness across the North American deserts.

I next addressed the hypothesis that adaptation to alternative host plant cytotypes drives reproductive isolation between herbivore populations. In Chapter 2, my investigation focused on two ecologically similar herbivores, the creosote bush grasshopper (*Bootettix argentatus*) and the creosote bush katydid (*Insara covilleae*). I first used RADcap sequencing to genotype hundreds of individuals from across the full distribution of each species. I then inferred range-wide population

structure and modeled spatial variation in gene flow across the species' range. To discern the relative role of ecology vs. geography in structuring insect populations, I then compared regions of reduced gene flow to physical barriers as well as creosote bush contact zones. My findings differed markedly between species. *Bootettix* comprises three major lineages, each associated with a different creosote bush cytotype. Two *Bootettix* lineages meet in a short hybrid zone that is coincident with the contact zone between tetraploid and hexaploid creosote bush, and gene flow was more restricted at this contact zone that across any physical barriers (e.g., mountain ranges, rivers). In contrast, *Insara* shows minimal population structure across most of its distribution. Although I found moderate subdivision across the Madrean Archipelago, evidence for host-associated genetic divergence was weak. This work shows that there are not necessarily uniform effects of host plant polyploidy even on ecologically similar herbivores.

In Chapter 3, I consider the forces that maintain adaptive genetic and phenotypic variation within populations. Two discrete color morphs of the desert clicker grasshopper (*Ligurotettix coquilletti*) coexist across the species' range: a uniform morph with homogeneous color, and a banded morph with strong light-dark contrast. This patterning likely affects the efficacy of crypsis to avoid predation, raising the question of why polymorphism persists, and why the proportion of each morph varies across sites. One possibility lies in life history differences between sexes: while territorial males are exposed to predation only on creosote bush stems, females are most vulnerable while laying eggs on the ground. I hypothesized that selection may thus favor different cryptic phenotypes in each sex (sexually antagonistic selection), thereby maintaining crypsis polymorphism within populations. With quantitative analysis of crypsis and the visual environments used by male and female grasshoppers, I found that different phenotypes are favored in each sex. I then used RADcap sequencing and association mapping to identify a single autosomal locus strongly associated with pattern polymorphism. Surprisingly, the divergent alleles at this locus are also found in a sister species, suggesting that sexually antagonistic selection may preserve ancient genetic variation.

For Christa, my partner,
and for Kiva Mae, my favorite naturalist.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Thanks first of all to Noah, who gave me the space to develop projects, to fail in a supportive environment, and to forge ahead in another direction. That freedom was a gift… and so were the nudges in the right direction when I got too far off course. Noah's enthusiasm, encouragement, and expansive vision of biology were a constant source of inspiration over the past seven years.

I am also grateful for the advice and guidance of my committee – Bree Rosenblum, Rauri Bowie, and Michael Nachman. They have helped me avoid missteps, disabused me of some magical thinking, and strengthened my work overall.

My science was influenced in some form or another by members of the Whiteman lab past and present. Thanks to the Arizona crew: Parris Humphrey, Andy Gloss, Ben Goldman-Huertas, Rick Lapoint, Jen Koop, Paul Nabity, Anna Nelson-Dittrich, Niels Groen, David Hembry. Thanks also to the Berkeley bunch: Kirsten Verster, Jessica Aguilar, Julianne Pelaez, Nicolas Alexandre, Hiromu Suzuki, Cathy Rushworth, Rebecca Duncan, Teruyuki Matsunaga, Marianna Karageorgi, and Carolina Reisenman. The Nachman lab has kept me honest in two states (and counting). Thanks in particular to Katya Mack, Noëlle Bittner, and Taichi Suzuki.

Across >10,000 miles of desert highway and >300 sampling localities I have had lots of help in the field. I thank Christa Kivarkis, Luke O'Connor, Nicolas Alexandre, Anthony Baniaga, Shea Lambert, Uri García Vásquez, Kelsey Yule, Jacob Hans, Myron Child, and Risa Takenaka for their considerable patience and hopper catching skills.

Collections in Mexico were made under SEMERNAT permit FAUT-0062, and I thank Tom van Devender, Ana Lilia Reina, Harry Brailovsky, and Marysol Trujano Ortega for assistance with permitting.

My last chapter benefitted from stops at two UC Reserves. I thank Chris Tracy (Boyd Deep Canyon Desert Research Center) and James André and Tasha La Doux (Sweeney Granite Mountains Desert Research Center) for their assistance.

The RADcap protocol I implemented in Chapters 2 and 3 was honed with advice from Lydia Smith and Ke Bi. Tips from Joyce Gross and Aaron Pomerantz improved my photography methodology for Chapter 3.

I am profoundly grateful to my undergraduate mentees – Jiarui Wang and Marissa Sandoval – whose dogged work on library preparation and image analysis enabled Chapter 3. You're both on to great things.

Thanks to colleagues for discussing the creosote bush system: Robert Laport, Jeff Joy, Tom van Devender, Camille Holmgren, and Julio Betancourt. Thanks to Kelsey Yule for collaboration, friendship, and overall desert enthusiasm.

For all their help with nuts and bolts of graduate student life, I thank Korshid Tarin, Emerita dela Cruz, and Monica Albe.

I wrote the final chapter at The Writers WorkSpace in Chicago. Thank you, Amy Davis, for creating such an effective place to work.

This work was only possible with the assistance of small grants from professional societies. These include the Society for the Study of Evolution Rosemary Grant Award, Society for Integrative and Comparative Biology Grant-In-Aid of Research, the Southwestern Association of Naturalists McCarley Research Grant, American Museum of Natural History Roosevelt Memorial Grant, Orthopterists' Society Cohn Research Grant, Essig Museum Walker Grant, American Society of Naturalists Student Research Award, and grants from the Department of Integrative Biology. Parts of this research were also funded by grants from the National Science Foundation (DEB-1405966) and National Institute of General Medical Sciences of the National Institutes of Health to Noah Whiteman (R35GM119816).

During my graduate work I was supported by a National Science Foundation Graduate Research Fellowship and Philomathia Graduate Fellowship in Environmental Sciences. The rest of the time I was teaching with fantastic colleagues who enriched my working life: Kelly Agnew, Alison Nguyen, and Erin Person.

Parris, Heather, and Larrea – good friends and desert rats – thank you for your support.

Sharu, Odie, Mom, and Dad carried me to the end.

Christa was along for the whole crazy ride. If I succeeded it was because of her.

CHAPTER 1: Polyploidy in creosote bush (*Larrea tridentata*) shapes the biogeography of specialist herbivores

This chapter has been previously published and is reproduced here with permission:

O'Connor T.K, Laport R.G., and Whiteman N.K. Polyploidy in creosote bush (*Larrea tridentata*) shapes the biogeography of specialist herbivores (2019) *Journal of Biogeography* **46**, 597-610.

ABSTRACT

Aim. Whole-genome duplication (polyploidy) can influence the biogeography and ecology of plants that differ in ploidy level (cytotype). Here, we address how two consequences of plant polyploidy (parapatry of cytotypes and altered species interactions) shape the biogeography of herbivorous insects.
Location. Warm deserts of North America.
Taxa. Gall midges (*Asphondylia auripila* group, Diptera: Cecidomyiidae) that attack three parapatric cytotypes of creosote bush (*Larrea tridentata*, Zygophyllaceae).
Methods. We surveyed *Asphondylia* species diversity at 177 sites across a 2300-km extent. After noting a correspondence between the distributions of eight *Asphondylia* species and *L. tridentata* cytotypes, we fine-mapped *Asphondylia* species range limits with transects spanning cytotype contact zones. We then tested whether plant-insect interactions and/or abiotic factors explain this coincidence by (1) comparing attack rates and gall midge communities on alternative cytotypes in a narrow zone of sympatry and (2) using species distribution models (SDMs) to determine if climatically suitable habitat for each midge species extended beyond cytotype contact zones
Results. The range limits of 6/17 *Asphondylia* species (including two novel putative species confirmed with *COI* sequencing) perfectly coincided with the contact zone of diploid and tetraploid *L. tridentata*. One midge species was restricted to diploid host plants while five were restricted to tetraploid and hexaploid host plants. Where diploid and tetraploid *L. tridentata* are sympatric, cytotype-restricted midge species more frequently attacked their typical host and *Asphondylia* community structure differed markedly between cytotypes. SDMs predicted that distributions of cytotype-restricted midge species were not constrained by climatic conditions near cytotype contact zones.
Main conclusions. Contact zones between plant cytotypes are dispersal barriers for many *Asphondylia* species due to plant-insect interactions. The distribution of *L. tridentata* cytotypes therefore shapes herbivore species ranges and herbivore community structure across North American deserts. Our results demonstrate that polyploidy in plants can affect the biogeography of ecological communities.

INTRODUCTION

Autopolyploidy (whole genome duplication without hybridization) is common in vascular plants and can influence plant ecology and evolution (Soltis et al., 2007; Parisod et al., 2010; Ramsey & Ramsey, 2014; Barker et al., 2015, 2016; Van de Peer et al., 2017). Related plants of different ploidy level (cytotype) are often reproductively isolated, which can limit local coexistence through reproductive interference (Levin, 1975) and drive rapid speciation (Coyne & Orr, 2004; Soltis et

al., 2007). Cytotypes may also differ in a suite of traits including size, growth form, phenology, water use, cold-hardiness, and secondary metabolism (Levin, 1983; Ramsey & Schemske, 2002) due to genome doubling *per se* (e.g., Stebbins, 1949) and/or subsequent evolutionary divergence (e.g., Ramsey, 2011). Accordingly, abiotic niches differ between cytotypes in some plant taxa (McIntyre, 2012; Thompson et al., 2014). However, broadly overlapping climatic niches in other taxa indicate that cytotypes may be in direct competition (Laport et al., 2013; Glennon et al., 2014). The joint effects of niche differentiation and population-level processes (reproductive interference and competition) likely underlie a common biogeographic consequence of polyploidy: cytotypes of many taxa have parapatric or allopatric distributions (Lewis, 1980).

The phenotypic effects of polyploidy can also influence interactions with other species, including herbivorous insects (Thompson et al., 1997, 2004; Segraves & Anneberg, 2016). In some systems, rates of herbivory are mediated by insect abundance in a cytotype's preferred microclimate (Arvanitis et al., 2007; Richardson & Hanks, 2011). Other insect herbivores consistently attack one host cytotype over others (Thompson et al., 1997; Nuismer & Thompson, 2001; Halverson et al., 2008a) or specialize exclusively on a single cytotype (Arvanitis et al., 2010). The herbivores that discriminate most strongly between host plant cytotypes tend to have highly specialized and intimate interactions with their host plants (e.g., gall-makers and other internal-feeding insects), which may make them especially sensitive to phenotypic differences among plant cytotypes (Segraves & Anneberg, 2016).

Given that polyploidy affects plant biogeography and species interactions, its influence on the biogeography of closely-associated species is a notable gap in our understanding of polyploidy's ecological significance. Thompson and colleagues (1997) hypothesized that the geographic distribution of favorable plant cytotypes may constrain the distribution of herbivore species. The distribution of *Greya politella* moths on *Heuchera grossulariifolia* did not support this hypothesis (Thompson et al., 1997). *H. grossulariifolia* comprises cytotypes with broadly overlapping distributions (Thompson et al., 1997; Segraves et al., 1999), as do most plants for which the effect of polyploidy on species interactions has been studied (Mandáková & Münzbergová, 2006; Arvanitis et al., 2007; Halverson et al., 2008b; Kao, 2008a) (but see Münzbergová et al., 2015). Plants with sympatric cytotypes allow for natural experiments to parse the effect of cytotype from those of geography and environment, but any effect of plant-insect interactions is not necessarily expected to shape the geographic range of herbivore species. Instead, we hypothesized that plants with parapatric cytotypes could constrain herbivore distributions if cytotype contact zones serve as biotically-mediated dispersal barriers.

To address this hypothesis, we tested if plant cytotype variation shapes the biogeography of herbivores of creosote bush (*Larrea tridentata* (DC.) Coville, Zygophyllaceae), a long-lived and dominant shrub in the warm deserts of North America. *Larrea tridentata* is a classical polyploid series comprising diploid ($2n = 2x = 26$), autotetraploid ($2n = 4x = 52$), and autohexaploid ($2n = 6x = 78$) cytotypes that roughly assort among the Chihuahuan, Sonoran, and Mojave Deserts (Barbour, 1969; Yang, 1970). Contact zones between *L. tridentata* cytotypes are well characterized in multiple areas, including sites where cytotypes are sympatric (Laport et al., 2012; Laport & Minckley, 2013; Laport & Ramsey, 2015). These resources enable comparisons with the distributions of insect herbivores and direct tests for the effect of cytotype on species interactions. Diploid and tetraploid cytotypes meet in a short and well-defined contact zone in southeastern

Arizona, while a sinuous contact zone between tetraploids and hexaploids extends from central Arizona to southern California (Laport et al., 2012). Established tetraploid populations are hypothesized to have a single origin (Laport et al., 2012). The precise timing of polyploidy events is unknown, but some of the oldest macrofossils of *L. tridentata* from packrat (*Neotoma* spp.) middens were inferred to be tetraploid (25,600 cal. y.b.p.; Cole, 1986; Hunter et al., 2001). Hexaploid populations are hypothesized to have arisen from tetraploids in the Holocene (Hunter et al., 2001; Holmgren et al., 2014) and may have multiple origins (Laport et al., 2012). Macrofossil evidence suggests that contemporary cytotype distributions were established in the past 5,000 years during expansion from glacial refugia (Hunter et al., 2001).

The creosote gall midges (*Asphondylia auripila* group, Diptera: Cecidomyiidae) are an adaptive radiation of 15 species that attack leaves, stems, buds, and flowers of *L. tridentata* (Gagné & Waring, 1990). Remarkably, the *A. auripila* group diversified without host plant switching (Joy & Crespi, 2007) and comprises the majority of the gall-forming insect community on *L. tridentata*. As with most gall midges (Gagné, 1989; Tokuda, 2012), the life cycles of *A. auripila* group are intimately associated with their host plant (Gagné & Waring, 1990). The small (2-5 mm) and short-lived adults emerge from their natal galls in one or two phenological windows per year. Many gall midges are weak fliers and actively disperse only short distances (i.e., a few meters, Highland, 1964), but wind-borne transport can enable passive dispersal over many kilometers (Yukawa et al., 2003; Yukawa & Rohfritsch, 2005; Miao et al., 2013). After mating, females deposit eggs and spores of a symbiotic fungus (Tokuda, 2012) into a species-specific host plant organ (Gagné & Waring, 1990). Larvae then induce the formation of an enclosed gall in which they develop to maturity. Although the ecology of the *A. auripila* group has been carefully studied at select sites (Waring & Price, 1989, 1990; Huggins, 2008), the distributions of *A. auripila* group species have not been previously characterized.

Our goal here was to test whether interactions with *L. tridentata* cytotypes constrained the distributions of individual *Asphondylia* species and shaped broader patterns of *Asphondylia* diversity across North American deserts. We had three specific aims: 1) determine the geographic distributions of species in the *A. auripila* group, 2) identify concordance between the range limits of *Asphondylia* species and *L. tridentata* cytotypes, 3) evaluate whether concordance is due to species interactions or confounding abiotic factors. In doing so, we link genome-scale mutational processes in plants to broad geographic patterns of herbivore diversity.

METHODS

Diversity and distributions of *Asphondylia auripila* group

We surveyed *Asphondylia* diversity on *L. tridentata* at 177 sites across the Chihuahuan, Sonoran, and Mojave Deserts between March 2015 and August 2016 (Fig. 1). At most sites we haphazardly collected branch segments from 5-10 plants, transported them to the laboratory, and a single investigator (TKO) searched them for galls with the naked eye. Two sites near Tucson, AZ were surveyed more intensively (up to 15 plants / site), and three sites were represented by opportunistic collections from a single plant. We sorted galls into morphotypes and identified the causative *Asphondylia* species using the original species descriptions (Gagné & Waring, 1990). Gall morphology is the most reliable diagnostic character for species of the *A. auripila* group (Gagné & Waring, 1990) and corresponds to species identity (Joy & Crespi, 2007).

Contact zone transects

To evaluate concordance between the range limits of *Asphondylia* species and *L. tridentata* cytotypes, we sampled along four transects spanning cytotype contact zones (Fig 1). We used two diploid–tetraploid transects (Gila and San Pedro) and two tetraploid–hexaploid transects (Salton and Bend) previously established by Laport and colleagues (Laport et al., 2012, 2016; Laport & Ramsey, 2015). Each transect includes 4–6 sites with permanently marked plants of known cytotype as determined by flow cytometry. In most transects the transition between parapatric cytotypes has been delimited to within ≤ 8 km (except Gila transect: 19 km). We did not sample marked plants directly to protect ongoing research. Instead, we sampled plants found among marked plants (< 5 m away) or at sites near those of Laport and colleagues. Additional details are provided in Appendix S1.

In all transects, sampling sites were ≥ 10 m from the nearest paved road and embedded in regionally typical vegetation. We collected ≥ 10 branch segments (30 cm long) from throughout the canopy of 10 focal plants per site. All plants were separated by at least 5 m to reduce the chance of sampling from the same clone and limit spatial autocorrelation among samples. Collections were stored in airtight containers under cool conditions for up to 24 hours. We then clipped stems into 5 cm-long segments, thoroughly mixed them, and subsampled 50 g fresh mass. A single investigator (TKO) exhaustively searched subsamples for galls with the naked eye. It was often impossible to differentiate senesced galls of the current season from those of previous seasons, which may remain on the plant for months. We included all galls in analyses described below, and results may therefore integrate patterns of gall midge diversity over several generations.

We tested whether the prevalence of *Asphondylia* species differed between cytotypes in each transect using binomial generalized linear mixed models (GLMMs) with site as a random effect. We evaluated the significance of model terms with Wald *F*-test and applied a 5% false discovery rate (FDR) correction to *P*-values to account for multiple testing. We only fit models for 13 species found on > 5 plants in a given transect. Analyses were performed in the R Environment for Statistical Computing v. 3.4.2 (R Core Team, 2017) with packages 'lme4' (Bates et al., 2015) and 'car' (Fox & Weisberg, 2011).

Diploid–tetraploid sympatry

To isolate the effect of cytotype from confounding factors, we next tested whether gall midge attack rates differed between sympatric diploid and tetraploid *L. tridentata* in the San Pedro River Valley (Fig. 1). Only a single such site has been characterized along the diploid–tetraploid contact zone. We surveyed galls on 72 permanently-marked plants of known cytotype (36 diploid, 36 tetraploid; Laport & Ramsey, 2015; RGL, unpublished data) near the peak of spring gall maturation in March–April 2016. To protect ongoing research on focal plants, we used a non-destructive field survey. A single investigator (TKO) inspected each plant for eight minutes under a standardized search routine and recorded all galls encountered. Although most plants were observed while blind to cytotype, morphological differences between diploids and tetraploids at this site (Laport & Ramsey, 2015) precluded a truly blind experiment. We measured mean diameter for 58/72 focal plants.

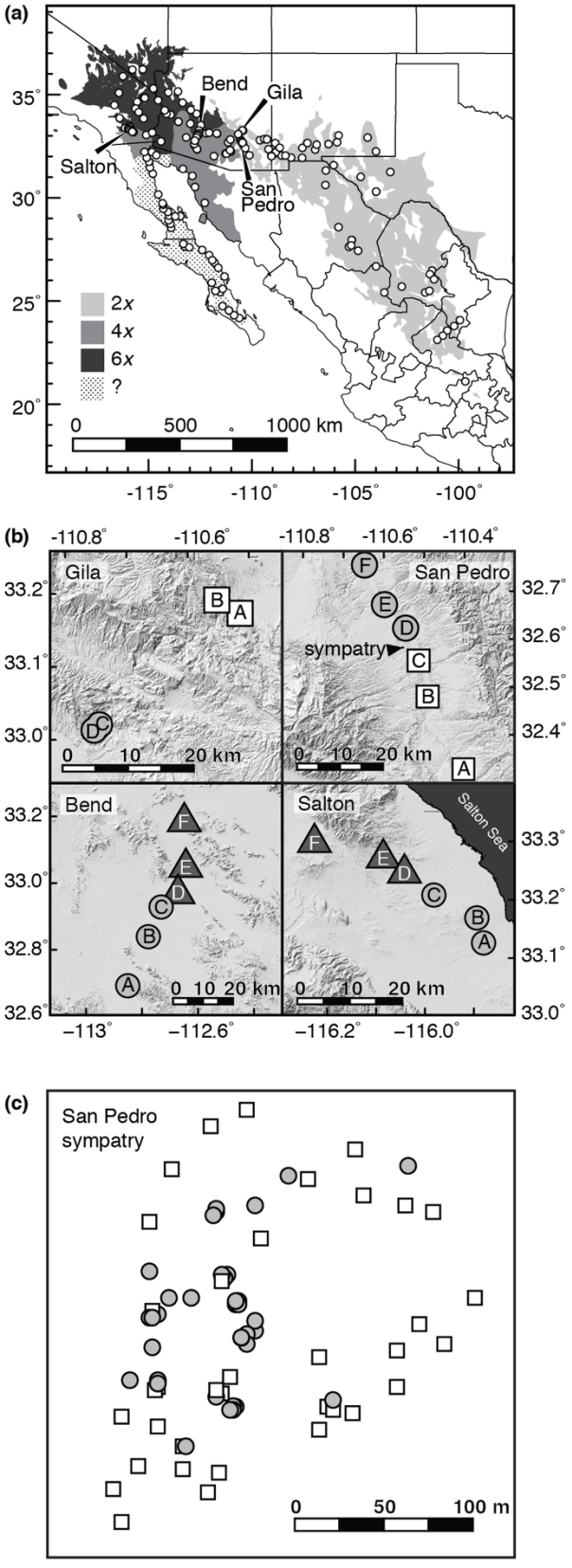

Figure 1. Map of sampling sites. **(a)** 177 total sites surveyed in this study. Shading shows the approximate distribution of diploid (2*x*), tetraploid (4*x*), and hexaploid (6*x*) *L. tridentata* (after Laport et al., 2015). *L. tridentata* cytotypes in the Lower Colorado River Valley and Baja California Peninsula have not been directly determined by flow cytometry. Arrowheads indicate transect locations. **(b)** Four contact zone transects. Symbols indicate creosote bush cytotype: white squares = 2*x*, gray circles = 4*x*, dark gray triangles = 6*x*. **(c)** Site of diploid–tetraploid sympatry within San Pedro transect showing location of 72 surveyed plants.

We used binomial generalized linear models (GLMs) to test whether the prevalence of *Asphondylia* galls differed between sympatric diploid and tetraploid hosts. We constructed GLMs for 13 *Asphondylia* species found on > 5 plants and assessed significance as for GLMMs. Initial models included cytotype, plant diameter, and their interaction as independent variables. Plant diameter did not differ between cytotypes (*t*-test, $P > 0.2$), so we present models with cytotype as the only independent variable. The inferred effect of cytotype from GLMs including or excluding plant diameter were generally consistent (Table S4.1). Analyses were performed with base functions in R and the 'car' package (Fox & Weisberg, 2011).

In addition to single-species analyses, we compared overall *Asphondylia* community structure between sympatric diploid and tetraploid *L. tridentata.* This analysis integrates over all *Asphondylia* species to test the hypothesis that interactions between *Asphondylia* species differ between *L. tridentata* cytotypes. We applied square-root and Wisconsin double-standardizations to gall abundance data (Bray & Curtis, 1957) and calculated Bray-Curtis dissimilarities among *Asphondylia* communities on individual plants. We analyzed only plants with ≥ 10 galls ($N = 54$). We then visualized variation among communities with non-metric multidimensional scaling (NMDS) ordinations and tested if community structure differed among cytotypes with permutational MANOVA (PERMANOVA) (Anderson, 2001). Analyses were implemented with the R packages 'MASS' (Venables & Ripley, 2002) and 'vegan' (Oksanen et al., 2017).

Species distribution models

As a complement to our field observations of *Asphondylia* distributions and abundance, we used species distribution models (SDMs) to assess whether abiotic factors or plant–insect interactions set the range limits of *Asphondylia* species. Our approach followed Anderson et al. (2002), who compared the predicted extent of climatically suitable habitat for parapatric spiny pocket mice (*Heteromys)* species to their observed distributions. Because *H. anomalus* was absent from climatically suitable areas that were occupied by *H. australis*, the authors concluded that competition set the range limit of *H. anomalus*. We conducted an analogous comparison for *Asphondylia* species with range limits near cytotype contact zones. If climatically suitable *Asphondylia* habitat extended beyond the cytotype contact zone, we would conclude that plant-insect interactions limit *Asphondylia* distributions. If suitable habitat was instead confined to the observed species distribution, we would conclude that abiotic factors are sufficient to explain the location of *Asphondylia* species range limits.

We built species distribution models (SDMs) using MaxEnt (Phillips et al., 2004). Models were fit using WorldClim v. 2.0 BioClim variables at 2.5 arc-minute resolution (Fick & Hijmans, 2017) and implemented with the R packages 'dismo' (Hijmans et al., 2017), 'ENMeval' (Muscarella et al., 2014), 'maxnet' (Phillips, 2017), 'MASS' (Venables & Ripley, 2002), 'raster' (Hijmans & van Etten, 2016), and 'spThin' (Aiello-Lammens et al., 2015). Expanded methods are reported in Appendix S2.

We selected eight uncorrelated BioClim variables ($|r| < 0.7$; Dormann et al., 2013) for modeling: Bio2 (mean diurnal temperature range), Bio 4 (temperature seasonality), Bio10 (mean temperature of warmest quarter), Bio11 (mean temperature of coldest quarter), Bio15 (precipitation seasonality), Bio16 (precipitation of wettest quarter), Bio17 (precipitation of driest quarter), and Bio19 (precipitation of coldest quarter). We combined our collection data with published surveys

of the creosote gall midge community (Werner & Olsen, 1973; Waring & Price, 1989; Gagné & Waring, 1990; Schowalter et al., 1999; Huggins, 2008).

Sampling bias can substantially affect Maxent model predictions (Kramer-Schadt et al., 2013), so we applied and compared two methods to account for bias in our dataset: spatial thinning of occurrence records (Pearson et al., 2007; Aiello-Lammens et al., 2015) and biased background sampling (Dudík et al., 2005). Spatially thinned analyses only included occurrences separated by ≥ 25 km, resulting in relatively uniform sampling across the range of each species. We randomly drew 10,000 background points from a species-specific background region (see below). The second set of SDMs used biased background sampling to match the bias in occurrence records. We modeled geographic variation in sampling intensity by estimating the 2-dimensional kernel density of our 177 sampling sites, then drew 10,000 background points with probability defined by local kernel density. To avoid model overfitting (Kremen et al., 2008), background regions for each species were defined as (1) within 100 km of the minimum convex polygon fit to occurrence records, and (2) within 100 km of the range of *L. tridentata*.

We tuned Maxent models by fitting 40 alternative combinations of feature classes and regularization multipliers and then selected the preferred model using the sample size-corrected Akaike information criterion (AICc) (Warren & Seifert, 2011). We fit final SDMs with Maxent v. 3.4 (Phillips et al., 2017) using optimized model settings for each species. We predicted habitat suitability across North America and estimated model performance using geographically-structured 4-fold cross-validation (Radosavljevic & Anderson, 2014; "block" method in 'ENMeval'). We converted probabilistic predictions of habitat suitability to binary predictions with the commonly-used 10% omission threshold (Pearson et al., 2007).

Evaluating SDM transferability

Our approach to test whether *L. tridentata* cytotypes determine the range limits of *Asphondylia* species assumes that SDMs can predict habitat suitability in unoccupied areas (i.e., that models are transferable across space). However, an array of methodological and biological factors can reduce model transferability (reviewed in Petitpierre et al., 2017). We addressed methodological challenges of inferring transferable SDMs using a limited number of uncorrelated, biologically meaningful predictors, as suggested by Petitpierre et al. (2017). Another persistent challenge is non-analog conditions between training and prediction regions (Fitzpatrick & Hargrove, 2009). The diploid–tetraploid contact zone coincides with the ecotone of the Chihuahuan and Sonoran Deserts, and the contrasting temperature and precipitation regimes in each desert region (MacMahon & Wagner, 1985) may limit SDM transferability.

We therefore evaluated SDM transferability across the study extent to identify potential limitations of our hypothesis testing framework. First, we sub-setted the occurrences of nine widespread species to include records only from the range of diploid *L. tridentata* ("diploid sites") or the range of polyploid *L. tridentata* ("polyploid sites"). These sub-setted records mimicked the observed distributions of some *Asphondylia* species (see Results). For each species, we then trained SDMs on either diploid or polyploid sites using model tuning and sampling bias correction as described above. We applied a 10% omission threshold calculated from training data (e.g., polyploid sites) and calculated omission rates for occurrences on the other side of the contact zone (e.g., diploid sites). Omission rates greater than the expected 10% would indicate limited model transferability

across desert regions. Finally, we compared habitat suitability predictions from models trained on complete vs. sub-setted occurrence records. By visualizing the agreement between these predictions across North American deserts, we identified where models trained on one region could be successfully transferred across the diploid–tetraploid contact zone and where they could not.

Molecular phylogenetics

Three gall morphotypes identified in this study did not match any species descriptions. We collected gall midges from two of these galls ("acuminata" and "hirsuta"; informal names explained in Appendix S3) and used molecular phylogenetics to test the hypothesis that they correspond to novel *Asphondylia* species. Detailed molecular methods and specimen information are reported in Appendix S3. Briefly, we sequenced fragments of *cytochrome c oxidase subunit I* (*COI*) from novel gall morphotypes ("acuminata", $N = 8$; "hirsuta", $N = 5$) and combined them with a previously ungenotyped member of the group (*A. discalis*, $N = 1$) and published sequences representing the other 14 described species in the *A. auripila* group (Joy & Crespi, 2007). Following initial phylogenetic analyses, we sequenced more exemplars of *A. florea* ($N = 10$), *A. apicata*, ($N = 3$), *A. rosetta* ($N = 3$) and *A. silicula*, ($N = 4$) to provide additional phylogenetic context. Our final dataset comprised 34 new *COI* sequences and 34 sequences from the literature representing all 15 described species in the *A. auripila* group, midges, two novel gall morphotypes, and seven outgroups (Table S3.2). We aligned sequences with MUSCLE (Edgar, 2004) using default parameters, trimmed sequences to a uniform length (432 bp), and estimated a maximum likelihood phylogeny with PHYML v. 3 (Guindon et al., 2010) under a GTR+G model. Branch support was estimated with 200 bootstrap replicates.

RESULTS

Diversity of the *A. auripila* group

We found 18 gall morphotypes from 177 total sites across a 2,300-km extent (Appendix S5). Among these were galls of all 15 species and three gall morphotypes that did not match any species description (Fig. S3.1). A large (10 mm-long) spade-shaped bud gall ("acuminata") was widespread in the Chihuahuan Desert. A smaller (2.5–3 mm), oblong gall found on the abaxial leaf surface ("hirsuta") resembled the gall of *A. silicula* but was densely covered in long trichomes. The "hirsuta" morphotype was also present in the Chihuahuan Desert but at lower frequency than "acuminata" galls. A third undescribed morphotype strongly resembled the gall of *A. silicula* but was attached to the adaxial leaf margin rather than the abaxial surface ("adaxial silicula"). The "adaxial silicula" galls were found in all contact zone transects, but because we did not record the position of *A. silicula* galls at most sites, the full distribution of this morphotype is unknown.

The *COI* phylogeny indicated that gall midges collected from "acuminata" and "hirsuta" galls represent distinct mitochondrial lineages in the *A. auripila* group (Fig. 2). *A.* sp. "acuminata" are monophyletic (bootstrap support [BS] = 99) and found as sister to *A. rosetta* (*A. apicata* + *A. florea*) with moderate support (BS = 88). *A.* sp. "hirsuta" was strongly supported as monophyletic (BS = 100), but relationships with other *Asphondylia* species were poorly resolved. We considered *A.* sp. "acuminata" and *A.* sp. "hirsuta" to be novel undescribed species based on the distinctness of their gall morphology, their phylogenetic placement, and the extent of mitochondrial sequence divergence from sister lineages (uncorrected divergence: *A.* sp. "acuminata" ≥ 4.3%, *A.* sp. "hirsuta" ≥ 7.7%).

We did not design our phylogenetic analyses to re-evaluate relationships within the *A. auripila* group or uncover cryptic diversity. Nevertheless, we note that *A. florea* specimens from across Arizona and California formed a clade (BS = 84) that was strongly supported as sister to the clade of *A. rosetta* (*A. apicata* + *A. florea*) sequenced by Joy & Crespi (2007) (BS = 93; Fig. 2). We also identified a divergent lineage of *A. silicula* from the southern Chihuahuan desert (represented by samples MX D8S4P2G1 and MX D7S2P2G1, Fig. S3.2) that renders the species paraphyletic with respect to *A. fabalis*. This lineage co-occurs with other *A. silicula* (MX D8S4P3G1) that are closely related to populations in the Sonoran Desert (*A. silicula* J&C1, Fig. S3.2). Additionally, we found that *A. discalis* (not sequenced by Joy & Crespi (2007)) is sister to *A. bullata* with strong support (BS = 100, Fig. S3.2).

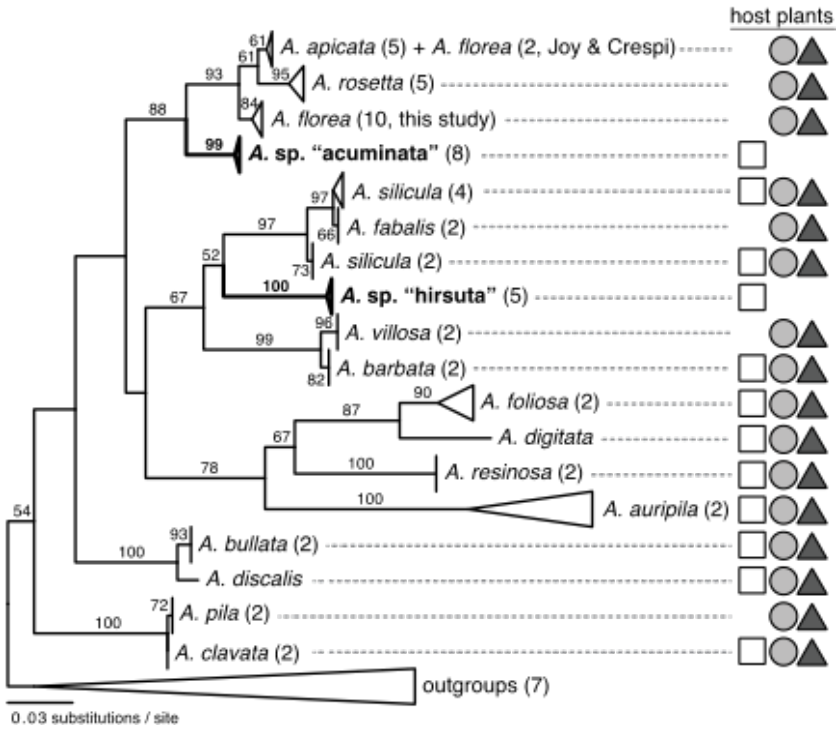

Figure 2. Maximum likelihood phylogeny of *Asphondylia auripila* group inferred from *cytochrome c oxidase subunit I* (*COI*) sequences. Two novel putative species (*A.* sp. "acuminata" and *A.* sp. "hirsuta") are indicated in bold. Number of samples per species given in parentheses. Bootstrap support values ≥ 50% are indicated above branches. Scale bar represents number of substitutions per site. Host plant use shown with symbols at right: white squares = 2*x*, gray circles = 4*x*, dark gray triangles = 6*x*.

Concordance of *Asphondylia* and *L. tridentata* range limits

Nine of 17 *Asphondylia* species were found across portions of all the Chihuahuan, Sonoran, and Mojave Deserts and therefore attack all *L. tridentata* cytotypes (Appendix S5). No species had range margins near the tetraploid–hexaploid contact zone. The other eight species were initially found only in the Chihuahuan Desert (on diploid hosts) or in the Sonoran and Mojave Deserts (on polyploid hosts). We found six such species in diploid–tetraploid transects, and each occurred on a single host cytotype. *A.* sp. "acuminata" was found in diploid sites while *A. apicata*, *A. florea*, *A. pila* (= "*A. pilosa*" of Gagne & Warning (1990) and Joy & Crespi (2007); see Gagné & Jaschof

(2014)), *A. rosetta*, and *A. villosa* were found in tetraploid sites (Fig. 3). "Adaxial silicula" galls were also found only in tetraploid sites. These species included all four bud-galling species, two leaf-galling species, and one leaf-galling morphotype (Table S5.1).

Concordance between *Asphondylia* range limits and the diploid–tetraploid contact zone is especially striking for *A. apicata*, *A. rosetta*, and "adaxial silicula" galls. Each was absent from diploid sites in the San Pedro transect but found at frequencies ≥ 0.5 in a tetraploid site < 8 km away (Fig. 3). We found similarly striking trends in the prevalence of *A. apicata, A. florea*, *A. rosetta*, and *A. silicula* in the Gila transect, although diploid and tetraploid sites were more distant than in the San Pedro transect. All six species and the "adaxial silicula" gall morphotype were also found in the sympatric site within the San Pedro transect. The range limits of these species were thus perfectly concordant with those of their typical host plant cytotype, a pattern we describe as "cytotype-restricted." Note that Joy & Crespi (2007) reported *A. florea* 65 km east of the diploid–tetraploid contact zone, within the range of diploid *L. tridentata*. In light of our phylogeny, this likely represents a different species from the *A. florea* we consider here.

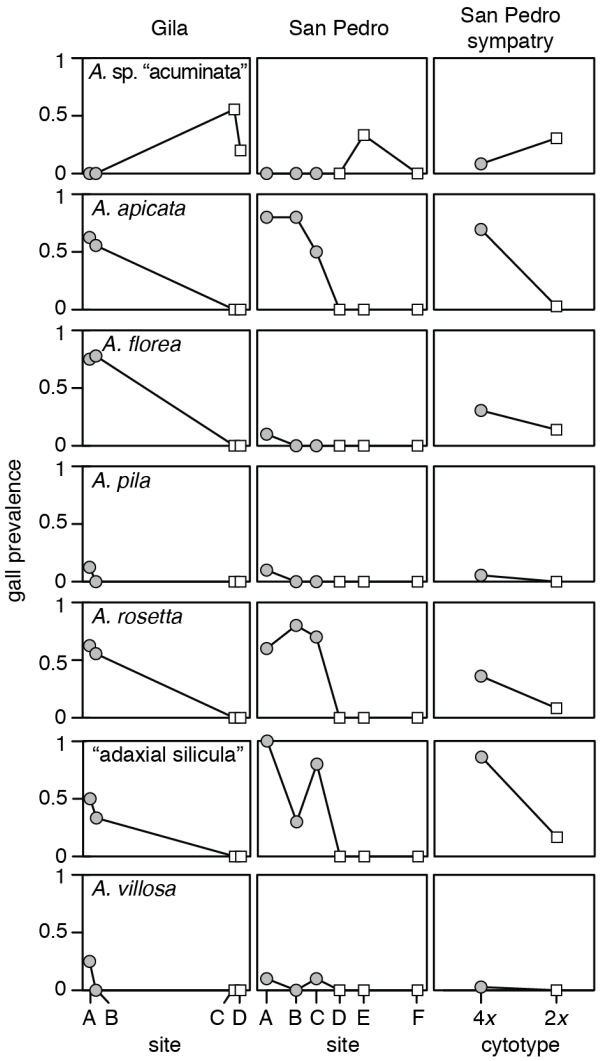

Figure 3. The range limits of six *Asphondylia* species and the adaxial *A. silicula* gall morphotype are concordant with the contact zone between diploid (white square) and tetraploid (gray circle) *L. tridentata*. For each species or morphotype, the first two columns show gall prevalence (proportion of plants with a gall) in contact zone transects. Relative position along the transect is shown on the *x*-axis. The third column shows gall prevalence on each plant cytotype where they naturally occur in sympatry.

Plant-insect interactions differ between *L. tridentata* cytotypes

All cytotype-restricted species attacked diploid and tetraploid hosts at different rates in sympatry. *A. pila* and *A. villosa* were rare in the sympatric site and found only on their typical tetraploid hosts ($N = 2$ and $N = 1$, respectively). *A.* sp. "acuminata", *A. apicata*, *A. rosetta*, and "adaxial silicula" galls were more prevalent on their typical host cytotype (binomial GLM, $P < 0.05$; Table 1). *A. florea* was also more than twice as prevalent on its typical host (11/36 tetraploids vs. 5/36 diploids), but this was only a trend ($P > 0.1$). Mean gall intensity (gall abundance on plants with ≥ 1 gall) for these four species and the "adaxial silicula" morphotype was higher on the species' typical host plant cytotype (1.3—3.3× higher). However, we had limited power to statistically evaluate differences between cytotypes because atypical cytotypes were rarely attacked.

Table 1. Summary of binomial generalized linear models and Wald *F*-tests for differences in prevalence of *Asphondylia* species between sympatric diploid (2*x*) and tetraploid (4*x*) *L. tridentata*. We fit models only for species found on ≥ 5 plants.

cytotype-restricted?	species	*N* 2*x*	*N* 4*x*	*b* (± s.e.)[a]	O.R. (95% CI)[a]	$F_{1,70}$[b]	*P*[b]
yes	*A.* sp. "acuminata"	11	3	-1.58 (± 0.70)	0.21 (0.04 – 0.74)	5.8	**0.035**
yes	*A. apicata*	1	25	4.38 (± 1.08)	79.5 (14.3 – 1504)	39.6	**<0.001**
yes	*A. florea*	5	11	1 (± 0.60)	2.73 (0.87 – 9.62)	2.87	0.137
yes	*A. pila*	0	2	-	-	-	-
yes	*A. rosetta*	3	13	1.83 (± 0.70)	6.22 (1.77 – 29.4)	8.3	**0.014**
yes	"adaxial silicula" morphotype	6	31	3.43 (± 0.66)	31 (9.3 – 125.8)	37.2	**<0.001**
yes	*A. villosa*	0	1	-	-	-	-
no	*A. auripila*	4	6	0.47 (± 0.69)	1.6 (0.42 – 6.78)	0.45	0.544
no	*A. barbata*	9	3	-1.30 (± 0.72)	0.27 (0.06 – 1.02)	3.64	0.099
no	*A. bullata*	15	15	0 (± 0.48)	1 (0.39 – 2.56)	0	1
no	*A. clavata*	17	5	-1.71 (± 0.59)	0.18 (0.05 – 0.54)	9.55	**0.009**
no	*A. digitata*	6	2	-1.22 (± 0.85)	0.29 (0.04 – 1.39)	2.28	0.176
no	*A. discalis*	3	28	3.65 (± 0.72)	38.5 (10.6 – 193)	38.5	**<0.001**
no	*A. foliosa*	0	2	-	-	-	-
no	*A. resinosa*	14	17	0.34 (± 0.48)	1.41 (0.55 – 3.63)	0.5	0.544
no	*A. silicula* (typical)	35	28	-2.30 (± 1.09)	0.10 (0.01 – 0.59)	6.78	**0.024**

[a] = Coefficients from GLM: *b*: estimated effect of cytotype, O.R.: odds ratio that galls of an *Asphondylia* species were found on a tetraploid rather than diploid plant. We report the inverse of the O.R. for *A.* sp. "acuminata" in the main text, which is the O.R. that this species was found on a diploid plant (its typical host) rather than a tetraploid.

[b] = Results of Wald *F*-tests. *P*-values reflect 5% false discovery rate correction.

Asphondylia community structure differed markedly between sympatric diploid and tetraploid *L. tridentata* (Fig. 4, PERMANOVA $R^2 = 0.33$, $P < 0.001$). This is consistent with the tendency of cytotype-restricted species to attack their typical host cytotype. However, the correlation of *Asphondylia* community structure and host cytotype was nearly as strong when cytotype-restricted species were removed from the dataset ($R^2 = 0.28$, $P < 0.001$). This indicates that some widespread *Asphondylia* species also attacked diploid and tetraploid hosts at different rates in sympatry. *A. discalis* was more prevalent on tetraploids in the sympatric site while *A. clavata*, *A. digitata*, and *A. silicula* were more prevalent on diploids (GLM, $P < 0.05$) (Table 1). The prevalence of *A.*

discalis was also higher in tetraploid sites than diploid sites within the San Pedro transect (diploid = 0.07, tetraploid = 0.66; GLMM, $P < 0.001$), although not in the Gila transect. By contrast, *A. clavata, A. digitata*, and *A. silicula* did not show consistent bias in prevalence in the San Pedro or Gila transects (Table S1.2).

In general, gall prevalence was similar on both cytotypes at the tetraploid–hexaploid contact zone. However, one species (*A. foliosa*) was more prevalent on tetraploids in the Salton transect (GLMM, $P < 0.05$, Table S1.2). We also detected differences in prevalence for several other species (*A. auripila*, *A. florea* in the Salton transect, *A. discalis* in the Bend transect), but these were not statistically different after FDR correction.

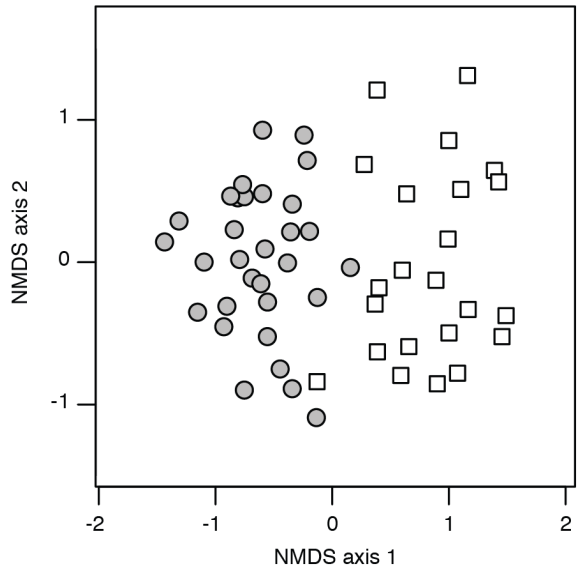

Figure 4. Non-metric multidimensional scaling (NMDS) ordination of *Asphondylia* communities on sympatric diploid (white square) and tetraploid (gray circle) plants. Cytotype was a significant predictor of community structure (PERMANOVA, R^2 = 0.33, $P < 0.001$)

Abiotic limits to *Asphondylia* species ranges

Accounting for sampling bias using a biased background yielded SDMs with visual fits that were generally better or comparable to those using spatial thinning. For simplicity we discuss only the results using biased background sampling, but we note where spatially thinned models differ.

Tuned SDMs for all cytotype-restricted species predicted suitable abiotic habitat beyond their observed range limit at the diploid–tetraploid contact zone (Fig. 5). Nearly the full range of *L. tridentata* was predicted to be suitable for *A. acuminata*, which is found only in the Chihuahuan Desert. Suitability predictions for *A. apicata*, *A. florea*, and *A. rosetta* included most of the Sonoran and Mojave Deserts and regions of unoccupied habitat in the Chihuahuan Desert. Unoccupied Chihuahuan Desert habitat was contiguous with the observed distribution of *A. rosetta*, but was interspersed with larger unsuitable regions for *A. apicata* and *A. florea*. The *A. villosa* SDM based on biased background selection produced the biologically unrealistic prediction of suitable habitat in the coastal plains of Texas and Tamaulipas. Bias may have been more difficult to correct in this case due to limited occurrence records ($N = 23$) that were spatially clustered. The SDM based on spatial thinning resembled those of *A. apicata*, *A. florea*, and *A. rosetta*, although suitable habitat in the Chihuahuan Desert was smaller than in those species. The SDM for *A. pila* least resembled other cytotype-restricted species. Regions beyond the northern limit of *L. tridentata* were predicted

to be suitable for *A. pila*, as were isolated patches of the southeastern Chihuahuan Desert. Further details of model parameters, evaluation, and performance are reported in Appendix S2.

Transferability experiments for nine widespread *Asphondylia* species revealed that models trained with polyploid sites systematically under-predicted suitable habitat in the Chihuahuan Desert, suggesting abiotically suitable habitat is more widespread than indicated by SDM predictions. Omission rate of diploid sites was well above the expected 10% for most species (up to 71%, Table 2), although *A. clavata* and *A. digitata* were exceptions to this trend. Under-prediction was even more pronounced in spatially thinned datasets (omission rates: 51-77%, Table S2.5). Notably, SDMs trained on polyploid sites predicted relatively small and non-contiguous regions of suitable habitat in the western Chihuahuan Desert (Fig. S2.1), similar to those observed in models for cytotype-restricted species (*A. apicata*, *A. florea*, *A. rosetta*, and *A. villosa*). By contrast, models trained only on occurrences from diploid sites tended to predict the distribution in the Sonoran and Mojave Deserts well (Table 2). The species with the least transferrable models (*A. auripila*, *A. bullata*, and *A. resinosa*) all have very limited distributions and / or sparse occurrence records confined to the Northern Chihuahuan Desert (Fig. S5.1). Species with broader ranges in the Chihuahuan Desert had omission rates near or below the expected 10% (range: 0-12%), indicating good model transferability. Results were qualitatively similar for spatially thinned models.

Table 2. Results of SDM transferability experiments with models employing biased background sampling. SDMs were trained on occurrences from the range of diploid or polyploid *L. tridentata*, then tested with occurrences from the opposite dataset. Columns report mean omission rates when applying a 10% omission threshold calculated from the training dataset.

Species	Training: polyploid Test: diploid	Training: diploid Test: polyploid
A. auripila	67%	38%
A. barbata	43%	3.30%
A. bullata	48%	57%
A. clavata	12%	2.80%
A. digitata	0%	8.30%
A. discalis	67%	5.20%
A. foliosa	71%	12%
A. resinosa	52%	45%
A. silicula	66%	5.90%

DISCUSSION

There is an emerging consensus that cytotype variation is an important dimension of plant biodiversity (Soltis et al., 2007; Ramsey & Ramsey, 2014; Barker et al., 2015; Laport & Ng, 2017), and growing interest in how polyploidy affects interactions with other species (Segraves & Anneberg, 2016). Only one study has formally tested whether host plant polyploidy influences the biogeography of herbivorous insects (Thompson et al., 1997). Here, we tested if cytotype variation in creosote bush (*L. tridentata*) affects interactions with specialized herbivores and shapes herbivore biogeography.

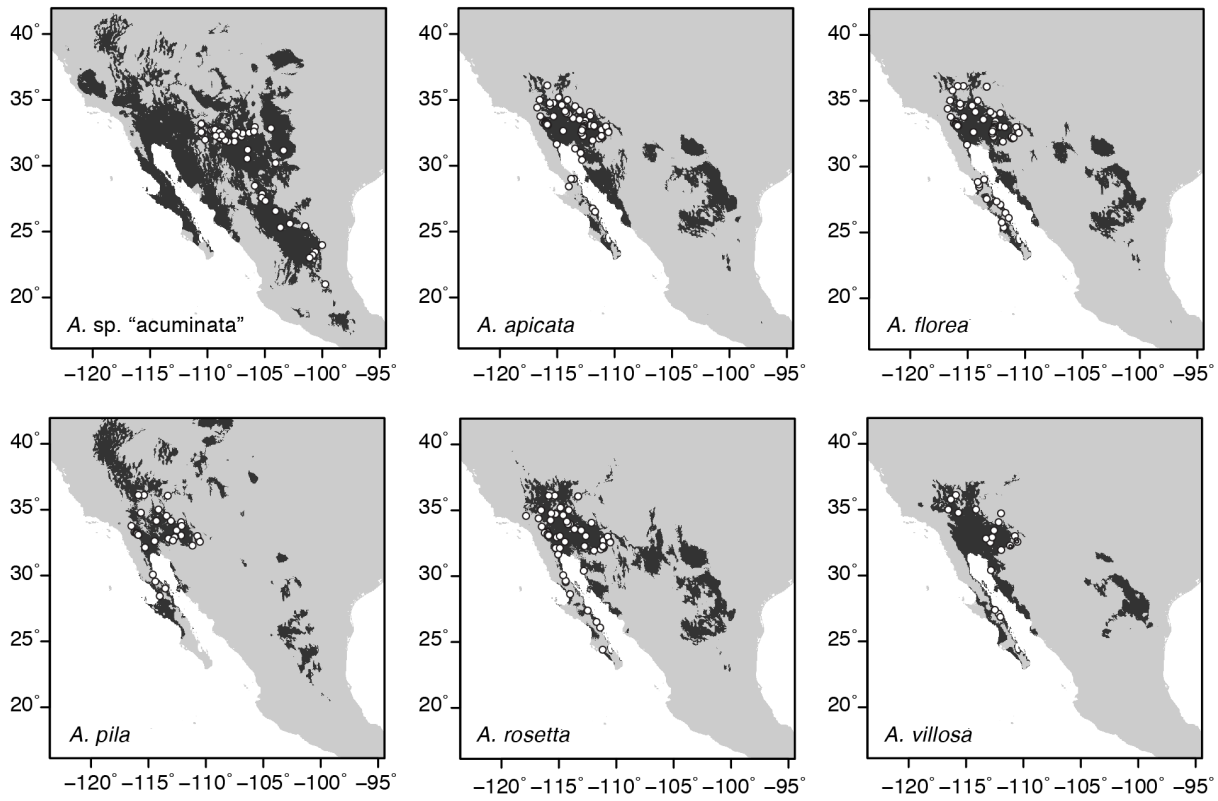

Figure 5. Species records and habitat suitability predictions for six cytotype-restricted species. Dark gray areas show suitable habitat when applying a 10% omission rate threshold. All species have predicted suitable habitat beyond their observed range limits. For species found on polyploid *L. tridentata* (all but *A.* sp. "acuminata"), tests of model transferability (Appendix S2) suggest species distribution models may under-predict the extent of suitable habitat in the Chihuahuan Desert. Predictions for all species except *A. villosa* are based on models fit using biased background sampling; *A. villosa* predictions are based upon predictions from spatially thinned models.

The observed range limits of six *Asphondylia* species and an additional gall morphotype coincided perfectly with the contact zone of diploid and tetraploid *L. tridentata*. Although it is likely that cytotype-restricted species disperse beyond their observed distributions, they must occur at only low densities: among > 5,800 galls identified from the San Pedro and Gila transects, cytotype-restricted species were only found on atypical host cytotypes where diploids and tetraploids are sympatric. Galls of all cytotype-restricted species were more prevalent on their typical host where diploid and tetraploid *L. tridentata* naturally co-occur, and overall *Asphondylia* community structure differed markedly between sympatric cytotypes. In contrast, we found only modest differentiation in *Asphondylia* galling between tetraploid and hexaploid *L. tridentata*. The same pool of *Asphondylia* species attacked tetraploids and hexaploids, and while several species tended to attack these cytotypes at different rates near contact zones, prevalence differed statistically for only one species in one transect. Although our findings are generally consistent across contact zone surveys, we note that these results would be more compelling with further replication.

Species distribution models (SDMs) indicated that abiotic factors are unlikely to determine the range limits of most cytotype-restricted species. Suitable habitat for *A.* sp. "acuminata" was

predicted across the full range of *L. tridentata*, while *A. apicata*, *A. florea*, *A. rosetta* were predicted to have large and nearly contiguous regions of suitable, but unoccupied, habitat in the Chihuahuan Desert. SDM transferability experiments demonstrated that models trained on occurrences from the range of polyploid creosote bush systematically underestimate suitable habitat in the northwest Chihuahuan Desert. Therefore, predicted distributions for *A. apicata*, *A. florea*, and *A. rosetta* on diploid *L. tridentata* are likely conservative. Both *A. pila* and *A. villosa* were predicted to have disjunct distributions separated by a broad region of unsuitable habitat in the western and central Chihuahuan Desert. It is unclear whether this reflects true abiotic tolerances or methodological bias, so we cannot exclude the possibility that abiotic and biotic factors jointly limit the distribution of these species.

It is possible that competition between *Asphondylia* species contributes to the parapatric distribution of some cytotype-restricted species (Sexton et al., 2009), although we have limited data to test this hypothesis. This is most plausible among bud-galling species, which include *A.* sp. "acuminata" on diploid hosts and *A. apicata*, *A. florea*, and *A. rosetta* on polyploids. We consider this unlikely because commonly find all three polyploid–restricted species on the same plant, some plants host all four species where diploid and tetraploid *L. tridentata* are sympatric, and many buds remain ungalled.

The *Larrea–Asphondylia* system demonstrates that extreme specialization on a host plant cytotype (as seen in *Dasineura cardaminis* on *Cardamine pratensis*, Arvanitis et al. 2010) is not required for polyploidy to shape insect biogeography. *Asphondylia* herbivory was biased between sympatric *L. tridentata* cytotypes, but less so than insect herbivory on *Solidago altissima* (Halverson et al., 2008a), *Heuchera grossulariifolia* (Nuismer & Thompson, 2001), and *Arnica cordifolia* (Kao, 2008b). In the *A. auripila* group, a host plant with parapatric cytotypes and modest differences in herbivore preference or performance appear sufficient to constrain herbivore distributions. Because cytotype parapatry is common among autopolyploid plants (Lewis, 1980), our results may be generalizable to other systems. For example, the sister lineage to the *A. auripila* group attacks a geographically segregated complex of *Atriplex canescens* cytotypes ($2x - 20x$) (Hawkins et al., 1986; Sanderson & Stutz, 1994) and is a compelling system for future study.

Several plausible mechanisms may underlie specialized interactions between *Asphondylia* species and host plant cytotypes. Traits that differ among *L. tridentata* cytotypes such as volatile organic compounds (Bohnstedt & Mabry, 1979) and organ size and proportions (Laport & Ramsey, 2015) affect host plant preference in other gall midges (Kanno & Harris, 2000; Hall et al., 2012). Alternatively, cytotype-restricted species may perform better on typical host cytotypes due to differences in plant defense. Although the major constituent of *L. tridentata* resin (nordihydroguaiaretic acid, NDGA) is a defense against many chewing herbivores (Mabry et al., 1977; Rhoades, 1977), internal-feeding *Asphondylia* spp. larvae are unlikely to encounter biologically meaningful levels of this compound. Moreover, NDGA levels are comparable between diploids and tetraploids (Zuravnsky, 2014). We are unaware of any studies investigating inducible defenses in *L. tridentata*, which are more likely to target internal parasites such as gall midges (e.g., Williams & Whitham, 1986; Harris et al., 2003).

Whether whole-genome duplication directly alters species interactions or potentiates their subsequent evolution is an unresolved question (Segraves & Anneberg, 2016). The limited

evidence from other systems is mixed. *Eurosta solidaginis* flies tend to attack independently-derived tetraploid plants over alternative cytotypes, suggesting that a trait directly affected by genome duplication *per se* might mediate plant-insect interactions (Halverson et al., 2008a). By contrast, although *Greya politella* moths prefer tetraploid *Heuchera grossulariifolia*, this is not due to pre-existing bias and likely evolved after polyploidization (Janz & Thompson, 2002). Comparisons between contact zone transects suggest an indirect effect of polyploidy in the *Larrea–Asphondylia* system. We found that while the prevalence of many *Asphondylia* species differed between adjacent diploids and tetraploid populations, few species attacked tetraploids and hexaploids at different rates. We hypothesize that this difference is due to 1) the more ancient divergence between tetraploids and diploids and 2) periods of allopatry between diploid and tetraploid *L. tridentata* during Pleistocene glacial cycles. However, we cannot rule out the possibility that phenotypic differences accompanying transitions from diploidy to tetraploidy present a steeper adaptive gradient for *Asphondylia* than transitions from tetraploidy to hexaploidy (Ramsey & Schemske, 2002; Madlung, 2013; Laport & Ramsey, 2015)

Joy & Crespi (2007) found that the *A. auripila* group diversified without host plant switching, a remarkable pattern among herbivorous insects. Our results suggest at least some diversification in this group may be associated with switching among *L. tridentata* cytotypes. For example, the diploid–restricted species *A.* sp. "acuminata" shares a common ancestor with the clade comprising *A. rosetta*, *A. apicata*, and *A. florea*, all of which are restricted to polyploid *L. tridentata* (Fig. 2). Robustly testing the hypothesis that speciation coincides with host cytotype switching will require greater phylogenetic resolution and elucidation of cryptic species in the *A. auripila* group.

Overall, our results are consistent with the hypothesis of Thompson and colleagues (1997) that polyploidy, by altering plant traits, can affect plant-insect interactions and impact herbivore distributions. The distribution of *L. tridentata* cytotypes indirectly contributes to the disparity in *A. auripila* group diversity between the Chihuahuan vs. Sonoran and Mojave Deserts (max 11 species / site vs. max 15 species / site, Fig S5.2). Given the strong influence of plant polyploidy on many species interactions (Segraves & Anneberg, 2016), genomic variation at the primary trophic level may affect the biogeography and composition of ecological communities. Future research on the community-wide effects of plant polyploidy (Laport & Ng, 2017; Segraves, 2017) should explicitly test for such biogeographic linkages.

DATA ACCESSIBILITY

COI sequences have been deposited in GenBank (accession numbers: MK058365-MK058398).

APPENDIX S1

Additional details of cytotype contact zone transects

METHODS

To evaluate concordance between the range limits of *Asphondylia* species and *L. tridentata* cytotypes, we sampled along four transects spanning cytotype contact zones: two 2*x*-4*x* transects (Gila and San Pedro) and two 4*x*-6*x* transects (Salton and Bend) that were previously established by Laport and colleagues (Laport et al., 2012, 2016; Laport & Ramsey, 2015). Abiotic and biotic features of the Gila, San Pedro, and Salton transects are described elsewhere (Laport & Minckley, 2013; Laport et al., 2016). Local cytotypes at each site were established by flow cytometry (see Laport et al., 2012 for general protocol). Each transect includes permanently marked plants of known ploidy. We did not sample these plants directly to protect ongoing research. Instead, we sampled plants found among marked plants (< 5 m) or at sites near those of Laport and colleagues. We did not verify the cytotype of study plants by flow cytometry. We denote sites from this study with letters to distinguish them from the numbering scheme of Laport & Ramsey. Comparisons between the transects of Laport and colleagues and this study are presented in Table S1.1.

The Gila transect in southeastern Arizona includes two diploid sites separated from two tetraploid sites by 19 km of precipitous river canyon. We sampled among or near (within 360 m) plants of known cytotype in April 2016.

The San Pedro transect in southeastern Arizona encompasses contiguous diploid and tetraploid populations in the San Pedro River Valley, where the complete transition between cytotypes has been localized to a 7.1 km region and includes a site of cytotype sympatry (Fig. 1c). In September 2015 we sampled at six sites along the San Pedro transect axis established by Laport & Ramsey (2015), but our sampling extended farther from the contact zone and included sites interspersed among those from the original study.

The Salton transect includes six sites that extend from the Salton Basin (-32 m) to the Santa Rosa Mountains (258 m) in southern California. The minimum distance between pure tetraploid and pure hexaploid sites was 7.4 km. We sampled among plants of known cytotype in April 2016.

The Bend transect (previously unreported) is centered on the town of Gila Bend in central Arizona, spanning 58 km over which there are no geographic barriers or obvious environmental transitions. Sites to the north and south of Gila Bend delimit the contact zone to a ≤ 8.1 km region. We sampled among plants of known cytotype in April 2016.

Table S1.1. Comparison between sites within contact zone transects in this study and those of Laport and colleagues. A dash indicates galls were surveyed at the same site reported by Laport and colleagues. Superscript indicates reference. [1] = Laport et al., 2012. [2] = Laport & Ramsey, 2015.

Site	Cytotype	Nearest site surveyed by flow cytometry	Distance to nearest site of same cytotype	Distance to nearest site of opposite cytotype	Notes
Gila A	2	Gila 4[2]	-	-	
Gila B	2	Gila 3[2]	360 m	18 km	
Gila C	4	Gila 2[2]	-	-	
Gila D	4	Gila 1[2]	-	-	
San Pedro A	2	AZ-U-2X[1]	8.1 km	32.3 km	between San Pedro A and nearest tetraploid site are four sites with only diploids
San Pedro B	2	AZ-07-T[1]	1.5 km	13.6 km	between San Pedro B and nearest tetraploid site are four sites with only diploids
San Pedro C	2	San Pedro 4[2]	110 m	5.1 km	San Pedro C is farther from the inferred contact zone than San Pedro 4
San Pedro D	4	San Pedro 2[2]	890 m	2.9 km	between San Pedro D and nearest diploid site there is one site with only tetraploids (n = 50 plants / site)
San Pedro E	4	AZ-10-S[1]	5.7 km	10.1 km	between San Pedro E and nearest diploid site there are two sites with on tetraploids (n = 9 and n = 50 plants / site)
San Pedro F	4	AZ-R[1]	730 m	20.4 km	
Bend A	4	AZ08-AQ[1]	-	-	
Bend B	4	AZ-08-AP[1]	-	-	
Bend C	4	AZ-08-AU[1]	-	-	
Bend D	6	AZ08-AT[1]	-	-	
Bend E	6	AZ07-V[1]	-	-	
Bend F	6	AZ08-AK[1]	100 m	30.7 km	
Salton A	4	CA-F-4x[1]	-	-	
Salton B	4	T2Q1-4x[1]	-	-	
Salton C	4	Salton 3[2]	-	-	
Salton D	6	Salton 4[2]	-	-	
Salton E	6	Salton 5[2]	-	-	
Salton F	6	CA09-Q[1]	-	-	

Table S1.2. Results of binomial generalized linear mixed models and Wald *F* tests for differences in *Asphondylia* spp. gall prevalence on *L. tridentata* cytotypes. We only fit models for species found on > 5 plant individuals in a given transect. GLMMs could not be fit for cytotype-restricted species in San Pedro and Gila transects. P_{FDR} represents *P* value after 5% false discovery rate (FDR) correction. Species with a significant effect of cytotype before FDR correction are bolded.

Transect	Species	*N* 2x	*N* 4x	*N* 6x	Wald *F*	d.f.	*P*	P_{FDR}
San Pedro	*A.* sp. "acuminata"	3/28	0/30	-	-	-	-	-
	A. apicata	0/28	21/30	-	-	-	-	-
	A. auripila	2/28	9/30	-	1.31	1, 55	0.258	0.775
	A. barbata	13/28	19/30	-	0.69	1, 55	0.410	0.798
	A. bullata	2/28	22/30	-	18.04	1, 55	**<0.001**	**<0.001**
	A. clavata	8/28	7/30	-	0.14	1, 55	0.707	0.929
	A. digitata	6/28	6/30	-	0.02	1, 55	0.894	0.975
	A. discalis	2/28	20/30	-	15.42	1, 55	**<0.001**	**<0.001**
	A. florea	0/28	1/30	-	-	-	-	
	A. foliosa	0/28	1/30	-	0.00	1, 55	0.975	0.975
	A. pila	0/28	1/30	-	-	-	-	
	A. resinosa	4/28	8/30	-	0.60	1, 55	0.444	0.798
	A. rosetta	0/28	21/30	-	-	-	-	
	A. silicula	24/28	24/30	-	0.13	1, 55	0.722	0.929
	"adaxial silicula"	0/28	21/30	-	-	-	-	-
	A. villosa	0/28	2/30	-	-	-	-	-
Gila	*A.* sp. "acuminata"	6/14	0/17	-	-	-	-	-
	A. apicata	0/14	10/17	-	-	-	-	-
	A. auripila	2/14	2/17	-	-	-	-	-
	A. barbata	10/14	12/17	-	0.00	1, 28	0.959	0.998
	A. bullata	7/14	1/17	-	5.15	1, 28	**0.031**	0.218
	A. clavata	10/14	8/17	-	1.83	1, 28	0.188	0.562
	A. digitata	7/14	1/17	-	1.44	1, 28	0.241	0.562
	A. discalis	10/14	12/17	-	0.00	1, 28	0.960	0.998
	A. florea	0/14	13/17	-	-	-	-	-
	A. pila	0/14	1/17	-	-	-	-	-
	A. resinosa	6/14	9/17	-	0.31	1, 28	0.581	0.998
	A. rosetta	0/14	10/17	-	-	-	-	-
	A. silicula	14/14	15/17	-	0.00	1, 28	0.998	0.998
	"adaxial silicula"	0/14	7/17	-	-	-	-	
	A. villosa	0/14	2/17	-	-	-	-	-
Bend	*A. apicata*	-	26/29	30/30	0	1, 56	1.000	1.000
	A. barbata	-	2/29	5/30	1.05	1, 56	0.310	0.569
	A. bullata	-	12/29	5/30	1.09	1, 56	0.302	0.569
	A. clavata	-	8/29	3/30	1.55	1, 56	0.218	0.569
	A. digitata	-	0/29	4/30	0.00	1, 56	1.000	1.000
	A. discalis	-	22/29	12/30	7.33	1, 56	**0.009**	0.099
	A. fabalis	-	1/29	0/30	-	-	-	-
	A. florea	-	12/29	11/30	0.02	1, 56	0.883	1.000
	A. foliosa	-	9/29	3/30	3.11	1, 56	0.083	0.306
	A. pila	-	0/29	3/30	-	-	-	-
	A. resinosa	-	14/29	22/30	3.78	1, 56	0.057	0.306

	A. rosetta	-	0/29	1/30	0.00	1, 56	0.996	1.000
	A. silicula	-	12/29	13/30	-	-	-	-
	"adaxial silicula"	-	2/29	0/30	0.01	1, 56	0.922	1.000

Salton	*A. apicata*	-	25/30	19/30	2.94	1, 57	0.092	0.220
	A. auripila	-	20/30	5/30	6.58	1, 57	**0.013**	0.068
	A. barbata	-	2/30	3/30	-	-	-	-
	A. bullata	-	3/30	2/30	-	-	-	-
	A. digitata	-	5/30	0/30	-	-	-	-
	A. discalis	-	17/30	20/30	0.08	1, 57	0.774	0.929
	A. fabalis	-	2/30	0/30	-	-	-	-
	A. florea	-	15/30	26/30	6.05	1, 57	**0.017**	0.068
	A. foliosa	-	27/30	4/30	17.46	1, 57	**<0.001**	**0.001**
	A. pila	-	17/30	10/30	0.87	1, 57	0.354	0.607
	A. resinosa	-	13/30	8/30	1.02	1, 57	0.317	0.607
	A. rosetta	-	3/30	15/30	3.30	1, 57	0.075	0.220
	A. silicula	-	1/30	2/30	-	-	-	-
	"adaxial silicula"	-	2/30	2/30	-	-	-	-

APPENDIX S2

Additional details of species distribution modeling

METHODS

To evaluate whether abiotic factors limit *Asphondylia* distributions, we built species distribution models (SDMs) using Maxent (Phillips et al., 2004). Models were fit using WorldClim v 2.0 BioClim variables at 2.5 arc-minute resolution (Fick & Hijmans, 2017) and implemented with the R packages 'dismo' (Hijmans et al., 2017), 'ENMeval' (Muscarella et al., 2014), 'maxnet' (Phillips, 2017), 'MASS' (Venables & Ripley, 2002), 'raster' (Hijmans, 2015), and 'spThin' (Aiello-Lammens et al., 2015).

Predictor selection

To choose uncorrelated predictor variables for species distribution modeling, we examined correlations among BioClim variables from 1,000 random points draw from the range of *L. tridentata*. We identified groups of correlated variables with hierarchical clustering and selected a single variable from each cluster such that remaining variables were uncorrelated ($|r| < 0.7$) (Dormann et al., 2013). When choosing among correlated variables we attempted to select those that were most proximal to the ecology of *Asphondylia* gall midges (Austin, 2007). Specifically, we preferred quarterly summaries of temperature and variables over monthly or yearly summaries, because *Asphondylia* life history appears to be tied to seasonal temperature and precipitation regimes that last several months at a time. Our final set of eight variables included Bio2 (mean diurnal temperature range), Bio 4 (temperature seasonality), Bio10 (mean temperature of warmest quarter), Bio11 (mean temperature of coldest quarter), Bio15 (precipitation seasonality), Bio16 (precipitation of wettest quarter), Bio17 (precipitation of driest quarter), and Bio19 (precipitation of coldest quarter). This approach the Petitpierre et al. (2017) to use a reduced set of environmental predictors related to species biology in order to produce models that are transferable across space.

Anderson (2017) recently proposed including information on host distribution when inferring SDMs of specialized parasites such as herbivorous insects. We chose not to include presence/absence of *L. tridentata* as a categorical variable in Maxent models because we focused on testing the hypothesis that abiotic factors (rather than biotic ones) explain the distribution of *Asphondylia* species. Models including *L. tridentata* presence produced qualitatively similar habitat suitability predictions results (not shown).

Occurrence records and background points

We combined our collection data with published surveys of the creosote gall midge community (Werner & Olsen, 1973; Waring & Price, 1989; Gagné & Waring, 1990; Schowalter et al., 1999; Huggins, 2008). The number of total records is reported in Table S2.1. Sampling bias can substantially affect Maxent model predictions (Kramer-Schadt et al., 2013). To address this, we applied and compared two methods to account for sampling bias in our dataset: spatial thinning of occurrence records (Pearson et al., 2007; Aiello-Lammens et al., 2015) and biased background sampling (Dudík et al., 2005). To avoid model overfitting (Kremen et al., 2008), background regions for each species were defined as (1) within 100 km of the minimum convex polygon fit to occurrence records, and (2) within 100 km of the range of *L. tridentata*.

Spatially thinned analyses only included occurrences separated by ≥ 25 km, resulting in relatively

uniform sampling across the range of each species. We randomly drew 10,000 background points from the species-specific background region. The second set of SDMs used biased background sampling to match the bias in occurrence records. We modeled geographic variation in sampling intensity by estimating the 2-dimensional kernel density (bandwidth = 1.5) of our 177 sampling sites, then drew 10,000 background points with probability defined by local kernel density.

Table S2.1. Number of occurrence records for all *Asphondylia* species. The number of thinned occurrences reflects the number of records retained after spatial thinning with a minimum distance of 25 km. *A. fabalis* and *A. foliosa* were not used for species distribution modeling.

species	total	thinned
A. sp. "acuminata"	44	37
A. apicata	73	42
A. auripila	94	74
A. barbata	83	58
A. bullata	82	56
A. clavata	91	66
A. digitata	60	41
A. discalis	101	67
A. fabalis	24	-
A. florea	63	41
A. foliosa	103	79
A. sp. "hirsuta"	19	-
A. pila	34	27
A. resinosa	82	49
A. rosetta	57	42
A. silicula	100	70
A. villosa	23	19

Model tuning

We identified optimal Maxent settings by comparing models fit with 40 alternative combinations of feature classes and regularization multipliers. All spatially thinned datasets had fewer than 80 occurrences. For both spatially thinned and biased background models we therefore allowed only combinations of linear (L), quadratic (Q), and hinge (H) features, as recommended by Maxent developers (Phillips & Dudík, 2008). Specifically, we fit models with L, H, LQ, LH, and LHQ feature classes. We explored eight regularization multipliers from 0.25-5 (0.25, 0.5, 1, 1.5, 2, 3, 4, 5). Alternative models were fit and compared using the R package 'ENMeval' (Muscarella et al., 2014). We evaluated models using geographically-structured 4-fold cross-validation (Radosavljevic & Anderson, 2014; "block" method in 'ENMeval') and summarized performance with area under the receiver operating curve (AUC), difference between training and test AUC (AUC_{diff}), as well as two threshold-based omission rates: OR_{MTP} (threshold = minimum value of a presence record in training set) or OR_{10} (threshold = value at which 10% of training data are omitted).

We then selected the preferred model for each species using the sample size-corrected Akaike information criterion (AICc) (Warren & Seifert, 2011). We fit final SDMs with Maxent v. 3.4 (Phillips et al., 2017) using optimized model settings for each species. We predicted habitat suitability across North America and again estimated model performance using geographically-structured 4-fold cross-validation. We converted probabilistic predictions of habitat suitability to

binary predictions using a 10% omission threshold (Pearson et al., 2007).

SDM transferability experiments

Using SDMs to test if abiotic factors limit the distributions of *Asphondylia* species requires that models have predictive power across desert regions (i.e., they are transferable). An array of methodological and biological factors can challenge model transferability (Petitpierre et al., 2017) including non-analog conditions between training and prediction regions (Fitzpatrick & Hargrove, 2009). The diploid–tetraploid contact zone coincides with the ecotone of the Chihuahuan and Sonoran Deserts, and the contrasting temperature and rainfall regimes to the east and west of the contact zone (MacMahon & Wagner, 1985) may limit SDM transferability. To assess this possibility, we fit SDMs for nine widespread *Asphondylia* species using only records from the range of diploid *L. tridentata* ("diploid sites") or only polyploid *L. tridentata* ("polyploid sites"). We coded all 193 sites as either diploid or polyploid based on their position relative to the diploid-tetraploid contact zone. Model tuning and bias correction follows methods described above. We then applied a 10% omission rate threshold calculated from training data (e.g., polyploid sites) and calculated omission rates for sampling points on the other side of the contact zone (e.g., diploid sites). Omission rates greater than the expected 10% would indicate limited model transferability. We also applied different omission rate thresholds used to classify habitat as suitable or unsuitable (0%, 5%, 20%) to determine whether our results were sensitive to threshold value.

RESULTS

Primary results presented and discussed in the main text.

Optimized model settings and performance

Settings and performance summaries for SDMs are presented in Table S2.2. Settings for models used in SDM transferability experiments are reported in Table S2.3 (models trained on polyploid sites) and Table S2.4 (models trained on diploid sites).

Transferability across desert regions

Results are discussed in main text. Difference between SDMs trained on all data vs. those trained on only polyploid sites is plotted in Fig S2.1. Note that for many species, models trained on polyploid sites predict much less contiguous habitat in the western Chihuahuan Desert than did models trained on all collection localities. (red regions near diploid-tetraploid contact zone). Omission rates in test data differed depending upon the threshold selected to classify habitat as suitable vs. unsuitable. However, our qualitative finding that models on diploid sites were more consistently transferable those trained on polyploid sites is unaffected by choice of threshold (Fig S2.2).

Table S2.2. Optimized Maxent settings for *Asphondylia* species distribution models. FC = feature classes (see text for details). RM = regularization multiplier. Mean and variance reflect estimates from 4-fold geographically-structured cross-validation. Akaike weight quantifies relative support for the preferred combination of FC and RM over alternative models.

				full dataset		cross-validation							
Bias correction method	**Species**	**FC**	**RM**	**Akaike weight**	**AUC**	**AUC**	**Var(AUC)**	**AUC_{diff}**	**Var(AUC_{diff})**	**OR_{10}**	**Var(OR_{10})**	**OR_{MTP}**	**Var(OR_{MTP})**
biased background	*A.* sp. "acuminata"	L	2	0.42	0.67	0.63	0.114	0.108	0.137	0.21	0.085	0.068	0.008
	A. apicata	LQ	1.5	0.4	0.73	0.72	0.009	0.024	0.002	0.04	0.007	0	0
	A. florea	LQ	1.5	0.81	0.72	0.65	0.059	0.099	0.067	0.27	0.11	0.188	0.057
	A. pila	LQ	0.5	0.93	0.78	0.72	0.071	0.093	0.08	0.21	0.087	0.122	0.031
	A. rosetta	LQ	0.3	0.91	0.67	0.58	0.133	0.134	0.186	0.3	0.219	0.161	0.076
	A. villosa	LQ	3	0.4	0.58	0.58	0.009	0.026	0.008	0.13	0.027	0.05	0.01
	A. auripila	LQ	1.5	0.66	0.65	0.63	0.05	0.076	0.029	0.22	0.044	0.053	0.004
	A. barbata	LQ	0.3	0.37	0.69	0.62	0.002	0.068	0.003	0.17	0.021	0.024	0.001
	A. bullata	LQ	0.5	0.62	0.7	0.67	0.062	0.091	0.02	0.17	0.013	0.012	0.001
	A. clavata	LQ	2	1	0.71	0.68	0.015	0.051	0.011	0.1	0.009	0.011	0.001
	A. digitata	LQ	0.3	1	0.7	0.62	0.073	0.133	0.04	0.18	0.016	0.067	0.009
	A. discalis	LQ	2	0.4	0.69	0.66	0.062	0.082	0.041	0.13	0.014	0.01	0
	A. foliosa	LQ	2	0.92	0.68	0.67	0.038	0.066	0.02	0.15	0.011	0.01	0
	A. resinosa	LQ	1.5	1	0.73	0.65	0.045	0.115	0.03	0.32	0.066	0	0
	A. silicula	LQ	1.5	0.79	0.68	0.6	0.126	0.13	0.067	0.32	0.095	0.04	0.001
spatial thinning	*A.* sp. "acuminata"	LQ	4	0.34	0.75	0.75	0.067	0.071	0.042	0.11	0.008	0.025	0.003
	A. apicata	LQ	1	0.97	0.79	0.77	0.024	0.043	0.009	0.17	0.021	0.023	0.002
	A. florea	LQ	0.5	0.78	0.81	0.81	0.021	0.038	0.017	0.15	0.09	0	0
	A. pila	LQ	0.3	0.87	0.83	0.7	0.074	0.149	0.07	0.36	0.116	0.107	0.019
	A. rosetta	LQ	0.5	0.7	0.75	0.73	0.061	0.084	0.045	0.34	0.186	0.068	0.019
	A. villosa	LQ	2	0.55	0.65	0.66	0.028	0.04	0.004	0.06	0.016	0	0
	A. auripila	LQH	2	0.54	0.78	0.78	0.033	0.046	0.026	0.13	0.046	0.083	0.028
	A. barbata	LQ	0.3	0.77	0.79	0.76	0.046	0.067	0.029	0.21	0.044	0.089	0.018
	A. bullata	LQ	0.3	0.58	0.8	0.75	0.052	0.107	0.027	0.29	0.095	0.179	0.056
	A. clavata	LQ	0.3	0.61	0.81	0.79	0.031	0.05	0.028	0.11	0.017	0.047	0.004
	A. digitata	LQ	0.3	0.75	0.79	0.72	0.043	0.113	0.019	0.24	0.043	0.118	0.02

	A. discalis	LQ	0.3	0.89	0.8	0.77	0.042	0.068	0.022	0.13	0.024	0.06	0.007
	A. foliosa	LQH	1.5	0.91	0.77	0.72	0.089	0.1	0.084	0.24	0.142	0.05	0.005
	A. resinosa	LQ	0.3	0.61	0.84	0.8	0.023	0.065	0.014	0.16	0.041	0.058	0.013
	A. silicula	LQ	0.5	0.78	0.78	0.7	0.065	0.112	0.017	0.22	0.021	0.043	0.003

Table S2.3. Optimized Maxent settings for *Asphondylia* species distribution models used in SDM transferability experiments. Models were trained and validated using occurrences from polyploid sites. Note that omission rates reported here summarize model performance using geographically-structured subsets of training localities. In the main text we summarize model performance on a fully independent set of test localities (diploid sites). See Table 2. Column definitions as in Table S2.2.

				full dataset		cross-validation							
Correction method	**species**	**FC**	**RM**	**Akaike weight**	**AUC (full)**	**AUC**	**Var(AUC)**	**AUC_{diff}**	**Var(AUC_{diff})**	**OR_{10}**	**Var(OR_{10})**	**OR_{MTP}**	**Var(OR_{MTP})**
biased background	*A. auripila*	LQ	0.25	0.37	0.651	0.61	0.126	0.119	0.103	0.3	0.156	0.18	0.106
	A. barbata	LQ	1	0.89	0.732	0.706	0.039	0.059	0.014	0.063	0.016	0	0
	A. bullata	LQ	0.5	0.84	0.74	0.706	0.021	0.051	0.018	0.155	0.022	0.031	0.004
	A. clavata	LQ	1	0.39	0.75	0.651	0.035	0.12	0.02	0.113	0.01	0.038	0.002
	A. digitata	LQ	0.5	0.82	0.685	0.593	0.136	0.167	0.119	0.438	0.215	0.25	0.199
	A. discalis	LQ	1	0.98	0.728	0.717	0.012	0.026	0.004	0.044	0.008	0	0
	A. foliosa	LQ	2	0.49	0.714	0.672	0.106	0.102	0.124	0.263	0.111	0.061	0.005
	A. resinosa	LQ	1	0.94	0.759	0.674	0.026	0.07	0.029	0.098	0.007	0	0
	A. silicula	LQ	0.5	0.99	0.752	0.731	0.06	0.065	0.049	0.161	0.056	0.054	0.005
spatial thinning	*A. auripila*	LQ	1	0.76	0.732	0.727	0.07	0.072	0.061	0.232	0.175	0.107	0.046
	A. barbata	LQ	1	0.52	0.769	0.742	0.009	0.011	0.001	0.036	0.005	0	0
	A. bullata	LQ	1	0.97	0.765	0.685	0.045	0.124	0.021	0.355	0.065	0.186	0.022
	A. clavata	LQ	0.5	0.88	0.802	0.806	0.034	0.042	0.021	0.028	0.003	0.028	0.003
	A. digitata	LQ	0.5	0.56	0.749	0.67	0.052	0.115	0.044	0.281	0.066	0.219	0.046
	A. discalis	LQ	0.5	0.93	0.792	0.802	0.025	0.034	0.006	0.05	0.003	0.025	0.003
	A. foliosa	LQ	0.5	0.68	0.765	0.742	0.058	0.075	0.051	0.265	0.12	0.045	0.008
	A. resinosa	LQ	1	0.95	0.81	0.774	0.049	0.047	0.014	0.17	0.02	0.056	0.012
	A. silicula	LQ	0.5	0.92	0.8	0.738	0.035	0.057	0.02	0.156	0.056	0	0

Table S2.4. Optimized Maxent settings for *Asphondylia* species distribution models used in SDM transferability experiments. Models were trained and validated using occurrences from diploid sites. Note that omission rates reported here summarize model performance using geographically-structured subsets of training localities. In the main text we summarize model performance on a fully independent set of test localities (polyploid sites). See Table 2. Column definitions as in Table S2.2.

				full dataset			cross-validation						
Correction method	**Species**	**FC**	**RM**	**Akaike weight**	**AUC (full)**	**AUC**	**Var(AUC)**	**AUC_{diff}**	**$Var(AUC_{diff})$**	**OR_{10}**	**$Var(OR_{10})$**	**OR_{MTP}**	**$Var(OR_{MTP})$**
biased background	*A. auripila*	LQH	2	0.74	0.82	0.77	0.02	0.058	0.01	0.24	0.038	0.05	0.01
	A. barbata	LH	2	0.91	0.72	0.56	0.06	0.19	0.08	0.36	0.135	0.083	0.003
	A. bullata	LQH	2	0.99	0.91	0.89	0.01	0.048	0.01	0.18	0.056	0.063	0.016
	A. clavata	LQ	5	0.26	0.71	0.64	0.11	0.118	0.13	0.23	0.148	0.128	0.036
	A. digitata	LH	2	0.39	0.91	0.87	0.02	0.034	0	0.08	0.028	0.083	0.028
	A. discalis	LQ	2	0.52	0.7	0.61	0.16	0.13	0.2	0.25	0.177	0.188	0.089
	A. foliosa	LQ	1	0.54	0.64	0.55	0.13	0.165	0.13	0.38	0.163	0.153	0.056
	A. resinosa	LQH	1	0.98	0.91	0.88	0.02	0.055	0.01	0.09	0.011	0.092	0.011
	A. silicula	LQ	1	0.65	0.69	0.57	0.07	0.163	0.07	0.23	0.033	0.085	0.005
spatial thinning	*A. auripila*	LQ	2	0.53	0.84	0.8	0.04	0.053	0.01	0.29	0.047	0	0
	A. barbata	L	4	0.2	0.76	0.71	0.01	0.03	0.01	0.11	0.019	0.036	0.005
	A. bullata	LQH	2	0.99	0.9	0.88	0.01	0.034	0	0.06	0.016	0.063	0.016
	A. clavata	LQH	4	0.32	0.8	0.76	0.14	0.104	0.13	0.25	0.168	0.179	0.073
	A. digitata	LH	2	0.18	0.9	0.87	0.01	0.017	0	0.29	0.118	0	0
	A. discalis	LQ	3	0.35	0.79	0.76	0.08	0.085	0.06	0.22	0.115	0.149	0.041
	A. foliosa	LQ	2	0.43	0.69	0.66	0.16	0.151	0.1	0.17	0.046	0.083	0.028
	A. resinosa	LQH	2	0.88	0.89	0.88	0.01	0.021	0	0.06	0.016	0.063	0.016
	A. silicula	LQ	3	0.38	0.74	0.69	0.16	0.126	0.18	0.22	0.141	0.028	0.003

Table S2.5. Results of SDM transferability experiments with models employing spatial thinning. SDMs were trained on occurrences from the range of diploid or polyploid *L. tridentata*, then tested with occurrences from the opposite dataset. Columns report mean omission rates when applying a 10% omission threshold calculated from the training dataset.

Species	**Training: polyploid Test: diploid**	**Training: diploid Test: polyploid**
A. auripila	75%	58%
A. barbata	51%	0.7%
A. bullata	74%	49%
A. clavata	72%	0%
A. digitata	56%	13%
A. discalis	75%	0%
A. foliosa	76%	0%
A. resinosa	76%	22%
A. silicula	77%	0%

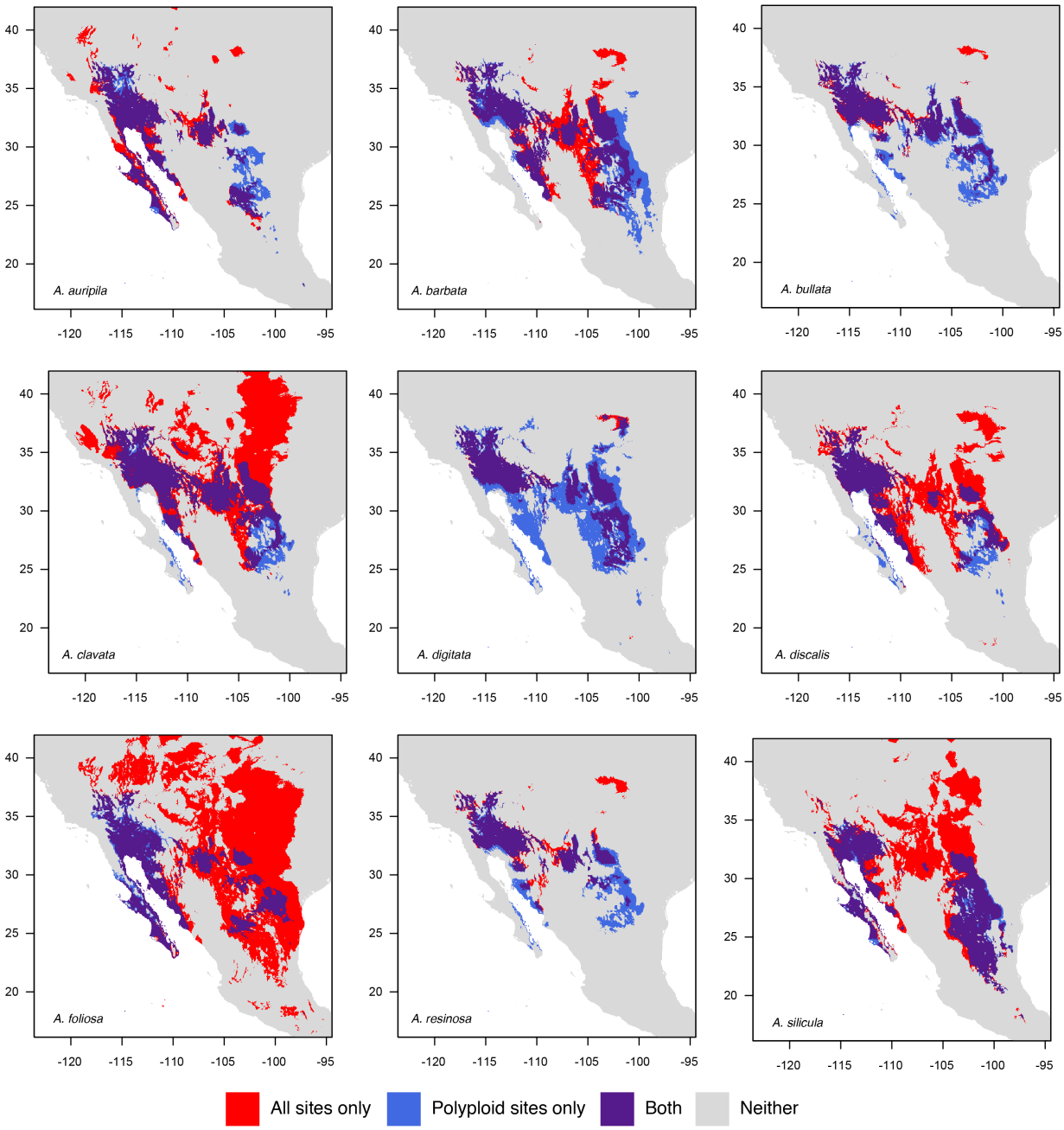

Figure S2.1. Comparison of species distribution models when trained on all sites or polyploid sites only. Sampling bias was corrected for using biased background sampling. Shaded regions indicate predicted suitable habitat when applying a 10% omission threshold to continuous Maxent predictions.

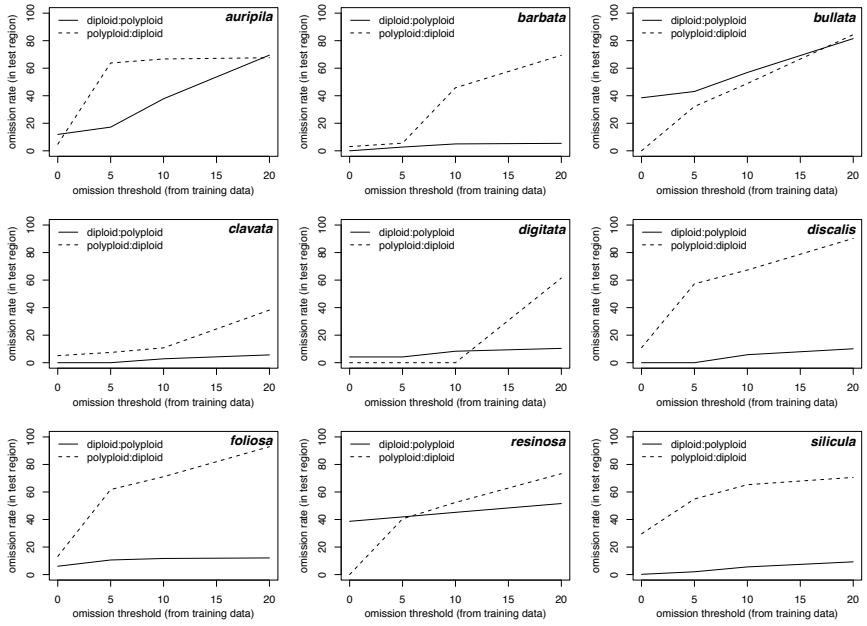

Figure S2.2. Comparison of omission rates in test regions (*y*-axis) when using alternative suitable/unsuitable threshold values calculated from training data (*x*-axis). For most focal species, models trained on polyploid sites and tested in diploid sites (polyploid:diploid) had higher omission rates than those trained on diploid sites and tested on polyploid sites (diploid:polyploid).

APPENDIX S3

Phylogenetic analysis of novel putative species

METHODS

Sampling Overview

Three gall morphotypes identified in this study did not match any published descriptions, suggesting they may be induced by novel *Asphondylia* species (Fig. S3.1). To test this hypothesis, we sequenced mitochondrial DNA fragments of gall midges collected from two of these undescribed gall morphotypes ("acuminata", $N = 8$; "hirsuta" $N = 6$) and combined our data with those of Joy & Crespi (2007). We also added exemplars of described species to provide additional phylogenetic context (*A. florea*, $N = 10$; *A. apicata*, $N = 3$; *A. rosetta*, $N = 3$; *A. silicula*, $N = 4$; *A. discalis*, $N = 1$). Collection details are provided in Table S3.1.

DNA extraction and sequencing

We extracted DNA from individual gall midges (adults, pupae, or late-instar larvae) with QIAGEN DNeasy Blood and Tissue kits according to manufacturer's instructions with the following modifications: samples were disrupted in a bead mill (20 Hz, 30 s) prior to an overnight proteinase K digestion at 56˚C, and DNA was eluted in two washes with 100 µL Buffer AE. We then amplified and bi-directionally sequenced a 472-bp fragment of cytochrome *c* oxidase subunit I (COI) from each specimen using primers C1-J-1718 and C1-N-2191 (Simon et al., 1994). A subset of samples was amplified with an alternative reverse primer (Asp-COI-R1: 5'-GTGTGTCAACTTCTATACCTACC), which we designed from the full-length COI sequence of *A. rosetta* (GenBank Accession: GQ387650). This alternative reverse primer targeted a slightly longer gene fragment (630 bp). Each 20 µl amplification reaction included 2 µL 10× PCR buffer (Thermo Scientific), 0.8 µl $MgCl_2$ at 50mM, 10 µL dNTPs at 10 mM, 0.4 µL of each primer at 10 µM, 0.1 µL recombinant Taq DNA polymerase (Thermo Scientific), 14.9 µL molecular-grade water, and 1 µL DNA. Cycling conditions consisted of an initial denaturation at 95˚C for 4 min; 35 cycles of 30 s denaturation at 95˚C, 30 s annealing at 55˚C, 45 s elongation at 72˚C; and a final elongation for 10 min at 72˚C. Reactions were cleaned and Sanger sequenced in both directions at the UC Berkeley DNA Sequencing Facility. Chromatograms were inspected for quality in GENEIOUS v. 10.2.3. (Biomatters, Ltd.), and forward and reverse reads were assembled into contigs. DNA sequences have been deposited in GenBank under accession numbers MK058365-MK058398 (Table S3.1).

Molecular phylogenetic analysis

We combined our data with COI sequences drawn from Joy & Crespi (2007) representing 14/15 described species in the *A. auripila* group and seven outgroups (Table S3.2). We aligned sequences with MUSCLE (Edgar, 2004) using default parameters, trimmed sequences to a uniform length (432 bp), and estimated a maximum likelihood phylogeny with PHYML v. 3 (Guindon et al., 2010) under a GTR+G model. Branch support was estimated with 200 bootstrap replicates.

RESULTS

Results are reported in main text. An expanded phylogeny showing individual samples is presented here (Fig. S3.2)

Table S3.1. Collection details and GenBank accession numbers for *Asphondylia* specimens sequenced for this study.

Species	Code	Country	State	County or Municipality	Latitude	Longitude	Date	Accession
A. sp. "acuminata"	acuminata 3/30/15 2	U.S.A.	Arizona	Graham	32.715	-109.698	3/30/15	MK058397
A. sp. "acuminata"	acuminata 3/30/15 17	U.S.A.	New Mexico	Doña Ana	32.498	-106.945	3/31/15	MK058396
A. sp. "acuminata"	acuminata 3/30/15 20	U.S.A.	Arizona	Graham	32.772	-109.487	3/30/15	MK058395
A. sp. "acuminata"	acuminata 3/30/15 35	U.S.A.	New Mexico	Hidalgo	32.083	-108.983	3/31/15	MK058394
A. sp. "acuminata"	acuminata 3/30/15 39	U.S.A.	New Mexico	Doña Ana	32.261	-107.016	3/30/15	MK058393
A. sp. "acuminata"	acuminata 3/30/15 41	U.S.A.	Arizona	Graham	32.715	-109.698	3/30/15	MK058392
A. sp. "acuminata"	acuminata 3/30/15 45	U.S.A.	Arizona	Graham	32.753	-109.354	3/30/15	MK058391
A. sp. "acuminata"	acuminata USC D3S1P4G1	U.S.A.	New Mexico	Eddy	32.843	-104.477	9/1/15	MK058390
A. apicata	apicata 3/23/15 26	U.S.A.	Arizona	Pinal	33.028	-111.770	3/23/15	MK058388
A. apicata	apicata 3/23/15 32	U.S.A.	Arizona	Mohave	34.535	-113.443	3/23/15	MK058387
A. apicata	apicata 3/23/15 8	U.S.A.	Arizona	Mohave	35.034	-114.133	3/23/15	MK058389
A. discalis	discalis 3/6/15 6	U.S.A.	Arizona	Pima	32.213	-110.999	3/6/15	MK058398
A. florea	florea 2/27/15 7	Mexico	Arizona	Pima	32.226	-111.137	2/27/15	MK058386
A. florea	florea 3/23/15 1	U.S.A.	Arizona	Yavapai	34.000	-112.763	3/23/15	MK058385
A. florea	florea 3/23/15 12	U.S.A.	California	Riverside	33.768	-115.333	3/24/15	MK058384
A. florea	florea 3/23/15 14	U.S.A.	California	San Bernardino	34.249	-115.723	3/24/15	MK058383
A. florea	florea 3/23/15 17	U.S.A.	Arizona	Yavapai	34.185	-113.044	3/23/15	MK058382
A. florea	florea 4/7/15 10	U.S.A.	Arizona	Pima	32.159	-111.107	4/7/15	MK058377
A. florea	florea 4/7/15 2	U.S.A.	Arizona	Pinal	33.077	-112.158	4/7/15	MK058381
A. florea	florea 4/7/15 4	U.S.A.	Arizona	Pinal	33.077	-112.158	4/7/15	MK058380
A. florea	florea 4/7/15 6	U.S.A.	Arizona	Maricopa	32.532	-112.881	4/7/15	MK058379
A. florea	florea 4/7/15 7	U.S.A.	Arizona	Pima	32.159	-111.107	4/7/15	MK058378
A. sp. "hirsuta"	hirsuta 3/30/15 10	U.S.A.	New Mexico	Doña Ana	32.261	-107.016	3/30/15	MK058376
A. sp. "hirsuta"	hirsuta 3/30/15 34	U.S.A.	New Mexico	Hidalgo	31.969	-108.642	3/31/15	MK058375
A. sp. "hirsuta"	hirsuta 3/30/15 75	U.S.A.	New Mexico	Hidalgo	31.969	-108.642	3/31/15	MK058374
A. sp. "hirsuta"	hirsuta USC D3S1P1G1	U.S.A.	New Mexico	Eddy	32.843	-104.477	9/1/15	MK058373
A. sp. "hirsuta"	hirsuta USC D3S1P4G7	U.S.A.	New Mexico	Eddy	32.843	-104.477	9/1/15	MK058372
A. rosetta	rosetta USM D1S7P3G1	U.S.A.	California	San Bernardino	34.652	-114.627	9/6/15	MK058371
A. rosetta	rosetta USM D1S7P7G1	U.S.A.	California	San Bernardino	34.652	-114.627	9/6/15	MK058370
A. rosetta	rosetta USM D1S8P1G1	U.S.A.	Nevada	Clark	35.222	-114.859	9/6/15	MK058369
A. silicula	silicula MX D5S2P5G1	Mexico	Coahuila	Castaños	26.392	-101.356	8/13/15	MK058368
A. silicula	silicula MX D7S2P2G1	Mexico	Nuevo Leon	Doctor Arroyo	23.705	-100.284	8/15/15	MK058367
A. silicula	silicula MX D8S4P2G1	Mexico	San Luis Potosí	Charcas	23.035	-101.100	8/16/15	MK058366
A. silicula	silicula MX D8S4P3G1	Mexico	San Luis Potosí	Charcas	23.035	-101.100	8/16/15	MK058365

Table S3.2. GenBank accession numbers and references for additional *Asphondylia* sequences included in molecular phylogeny.

Species	Code	Reference	Accession
A. apicata	*A. apicata* J&C1	Joy & Crespi (2007)	EF189965
A. apicata	*A. apicata* J&C2	Joy & Crespi (2007)	EF189966
A. rosetta	*A. rosetta* J&C1	Joy & Crespi (2007)	EF189967
A. rosetta	*A. rosetta* J&C2	Joy & Crespi (2007)	EF189968
A. florea	*A. florea* J&C1	Joy & Crespi (2007)	EF189969
A. florea	*A. florea* J&C2	Joy & Crespi (2007)	EF189970
A. auripila	*A. auripila* J&C1	Joy & Crespi (2007)	EF189973
A. auripila	*A. auripila* J&C2	Joy & Crespi (2007)	EF189974
A. foliosa	*A. foliosa* J&C1	Joy & Crespi (2007)	EF189971
A. foliosa	*A. foliosa* J&C2	Joy & Crespi (2007)	EF189972
A. resinosa	*A. resinosa* J&C1	Joy & Crespi (2007)	EF189975
A. resinosa	*A. resinosa* J&C2	Joy & Crespi (2007)	EF189976
A. barbata	*A. barbata* J&C1	Joy & Crespi (2007)	EF189977
A. barbata	*A. barbata* J&C2	Joy & Crespi (2007)	EF189978
A. clavata	*A. clavata* J&C1	Joy & Crespi (2007)	EF189979
A. clavata	*A. clavata* J&C2	Joy & Crespi (2007)	EF189980
A. fabalis	*A. fabalis* J&C1	Joy & Crespi (2007)	EF189985
A. fabalis	*A. fabalis* J&C2	Joy & Crespi (2007)	EF189986
A. pilosa	*A. pilosa* J&C1	Joy & Crespi (2007)	EF189981
A. pilosa	*A. pilosa* J&C2	Joy & Crespi (2007)	EF189982
A. silicula	*A. silicula* J&C1	Joy & Crespi (2007)	EF189987
A. silicula	*A. silicula* J&C2	Joy & Crespi (2007)	EF189988
A. villosa	*A. villosa* J&C1	Joy & Crespi (2007)	EF189983
A. villosa	*A. villosa* J&C2	Joy & Crespi (2007)	EF189984
A. digitata	*A. digitata* J&C1	Joy & Crespi (2007)	EF189989
A. bullata	*A. bullata* J&C1	Joy & Crespi (2007)	EF189990
A. bullata	*A. bullata* J&C2	Joy & Crespi (2007)	EF189991
A. caudicis	*A. caudicis*	Joy & Crespi (2007)	EF189992
A. atriplicis	*A. atriplicis*	Joy & Crespi (2007)	EF189993
A. neomexicana	*A. neomexicana*	Joy & Crespi (2007)	EF189994
A. bigeloviabrassicoides	*A. bigeloviabrassicoides*	Joy & Crespi (2007)	EF189995
A. sp.	*A.* sp.	Joy & Crespi (2007)	EF189996
A. websteri	*A. websteri*	Joy & Crespi (2007)	EF189997
A. conglomerata	*A. conglomerata*	Uechi et al. (2004)	AB115566

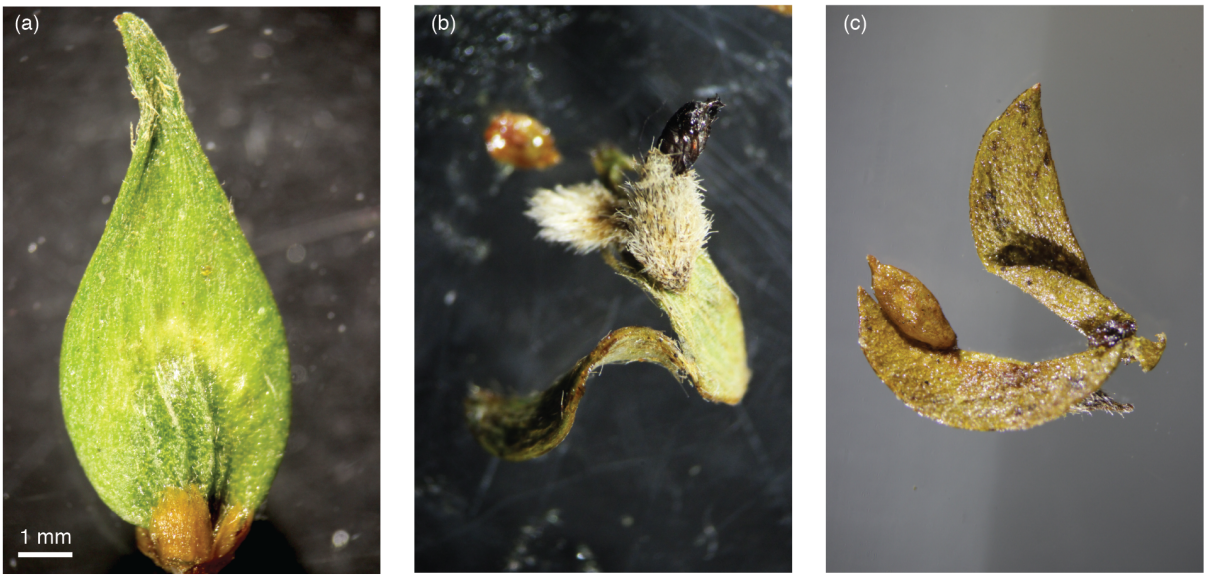

Figure S3.1. Exemplars of undescribed gall morphotypes identified in this study, which may be induced correspond to novel *Asphondylia* species. All images are on a common scale. (a) Freshly-collected gall of *A.* sp. “acuminata”. The provisional name for this putative species describes the gall’s spade-like (acuminate) morphology. (b) Dried gall of *A.* sp. “hirsuta” with midge puparium emerging in upper right corner. The provisional name for this putative species reflects the dense (hirsute) trichome growth on the gall’s surface. Trichome density is variable among individuals, however, and immature galls of this species can be difficult to distinguish from *A. silicula*. (c) Dried gall similar to that of *A. silicula* attached to inner margin of upper (adaxial) leaf surface. Typical *A. silicula* galls attach on the lower (abaxial) surface. We were unable to collect midges from this gall morphotype, so it is unclear whether it is induced by *A. silicula* or a novel species.

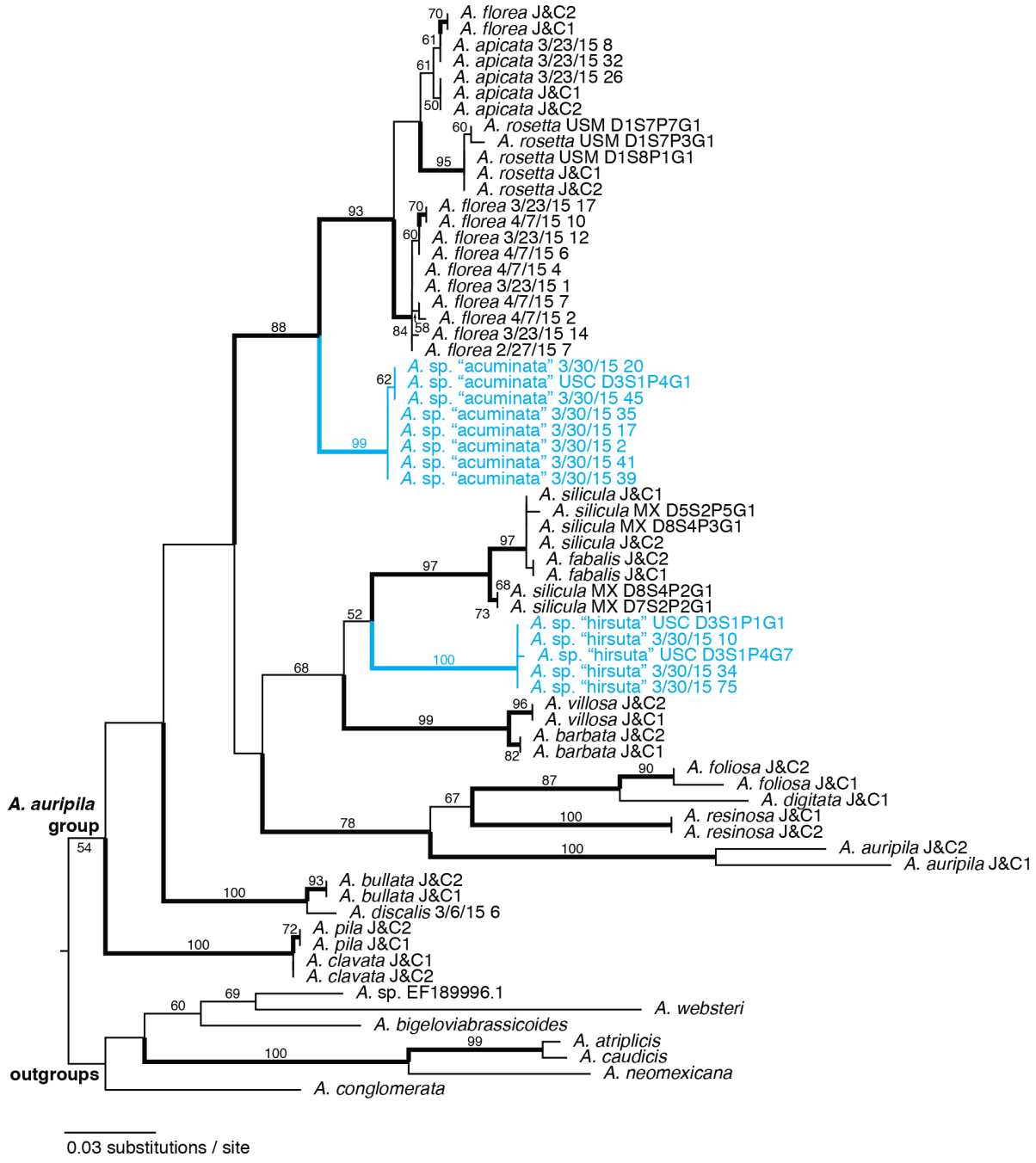

Figure S3.2. Maximum likelihood phylogeny of *Asphondylia auripila* group inferred from cytochrome *c* oxidase subunit I (COI) sequences. Branches with ≥ 70% bootstrap support are thickened and branches with ≥ 50% support are labeled. Two novel putative species (*A.* sp. "acuminata" and *A.* sp. "hirsuta") are indicated in blue.

APPENDIX S4

Supplementary results of generalized linear models

Table S4.1. Comparison of nested generalized linear models and Wald *F*-tests for differences in *Asphondylia* spp. prevalence between sympatric diploid (2*x*) and tetraploid (4*x*) creosote bush. We fit models only for species found on ≥ 5 plants. The effect of cytotype on *Asphondylia* spp. prevalence does not depend on whether plant size is included in the model of cytotype-restricted species (indicated with asterisk).

		Full model[a]				Main effects only[a]				Cytotype only[a]				Cytotype only[b]			
	variable	**SS**	$F_{1, 54}$	***P***	***P* (FDR)[c]**	**SS**	$F_{1, 55}$	***P***	***P* (FDR)[c]**	**SS**	$F_{1, 56}$	***P***	***P* (FDR)[c]**	**SS**	$F_{1, 70}$	***P***	***P* (FDR)[c]**
A. sp. "acuminata"*	cytotype	4.77	5.67	**0.021**	**0.034**	4.77	5.03	**0.029**	0.194	5.29	5.10	**0.028**	**0.045**	5.97	5.80	**0.019**	**0.035**
	diameter	0.90	1.07	0.305	0.441	0.90	0.95	0.334	1.000								
	cytotype:diameter	2.67	3.17	0.080	0.349												
	residuals	45.4				52.1				58.0				72.0			
*A. apicata**	cytotype	35.1	37.55	**<0.001**	**<0.001**	35.1	32.36	**<0.001**	**<0.001**	35.5	34.31	**<0.001**	**<0.001**	40.7	39.60	**<0.001**	**<0.001**
	diameter	0.10	0.11	0.746	0.746	0.10	0.09	0.763	1.000								
	cytotype:diameter	0.69	0.74	0.394	0.731												
	residuals	50.5				59.7				58.0				72.0			
A. auripila	cytotype	0.25	0.22	0.639	0.678	0.25	0.23	0.637	1.000	1.17	1.13	0.293	0.317	0.47	0.45	0.503	0.544
	diameter	4.61	4.08	**0.048**	0.187	4.61	4.14	**0.047**	0.607								
	cytotype:diameter	0.00	0.00	0.955	0.983												
	residuals	61.0				61.3				58.0				72.0			
A. barbata	cytotype	3.53	3.53	0.066	0.085	3.53	3.59	0.064	0.254	2.52	2.43	0.125	0.162	3.74	3.64	0.061	0.099
	diameter	2.00	2.00	0.163	0.273	2.00	2.03	0.160	1.000								
	cytotype:diameter	0.27	0.27	0.607	0.983												
	residuals	53.9				54.1				58.0				72.0			
A. bullata	cytotype	0.20	0.17	0.678	0.678	0.20	0.19	0.666	1.000	0.59	0.57	0.455	0.455	0.00	0.00	1.000	1.000
	diameter	4.31	3.77	0.057	0.187	4.31	4.06	**0.049**	0.607								
	cytotype:diameter	1.71	1.50	0.227	0.589												
	residuals	61.8				58.4				58.0				72.0			
A. clavata	cytotype	11.0	10.24	**0.002**	**0.006**	11.0	9.82	**0.003**	**0.028**	9.37	9.05	**0.004**	**0.010**	9.83	9.55	**0.003**	**0.009**
	diameter	2.24	2.08	0.155	0.273	2.24	1.99	0.164	1.000								
	cytotype:diameter	1.94	1.80	0.185	0.589												
	residuals	58.2				61.8				58.0				72.0			
A. digitata	cytotype	4.17	7.04	**0.010**	**0.019**	4.17	5.11	**0.028**	0.194	3.36	3.24	0.077	0.111	2.34	2.28	0.136	0.176
	diameter	1.16	1.95	0.168	0.273	1.16	1.42	0.239	1.000								
	cytotype:diameter	7.35	12.41	**0.001**	**0.006**												
	residuals	32.0				44.9				58.0				72.0			

A. discalis	cytotype	28.5	26.80	**<0.001**	**<0.001**	28.5	27.19	**<0.001**	**<0.001**	29.9	28.87	**<0.001**	**<0.001**	39.6	38.5	**<0.001**	**<0.001**
	diameter	0.34	0.32	0.576	0.624	0.34	0.32	0.573	1.000								
	cytotype:diameter	0.02	0.02	0.899	0.983												
	residuals	57.3				57.6				58.0				72.0			
*A. florea**	cytotype	4.46	4.19	**0.045**	0.066	4.5	4.28	**0.043**	0.216	5.35	5.16	**0.027**	**0.045**	2.95	2.87	0.095	0.137
	diameter	0.89	0.83	0.365	0.444	0.89	0.85	0.360	1.000								
	cytotype:diameter	0.04	0.03	0.854	0.983												
	residuals	57.5				57.3				58.0				72.0			
A. resinosa	cytotype	0.86	0.80	0.376	0.445	0.86	0.81	0.372	1.000	1.55	1.50	0.226	0.267	0.51	0.50	0.484	0.544
	diameter	4.28	3.95	0.052	0.187	4.28	4.02	**0.050**	0.607								
	cytotype:diameter	1.15	1.06	0.307	0.666												
	residuals	58.6				58.6				58.0				72.0			
*A. rosetta**	cytotype	8.58	8.58	**0.005**	**0.011**	8.58	8.65	**0.005**	**0.043**	10.5	10.14	**0.002**	**0.008**	8.53	8.30	**0.005**	**0.014**
	diameter	2.98	2.98	0.090	0.234	2.98	3.00	0.089	0.798								
	cytotype:diameter	0.06	0.06	0.803	0.983												
	residuals	54.0				54.6				58.0				72.0			
A. silicula (typical)	cytotype	6.60	13.76	**<0.001**	**0.002**	6.60	8.27	**0.006**	**0.046**	5.73	5.53	**0.022**	**0.045**	6.98	6.78	**0.011**	**0.024**
	diameter	3.05	6.36	**0.015**	0.187	3.05	3.82	0.056	0.607								
	cytotype:diameter	7.43	15.49	**<0.001**	**0.003**												
	residuals	25.9				43.9				58.0				72.0			
A. silicula (adaxial)*	cytotype	34.7	33.36	**<0.001**	**<0.001**	34.7	34.00	**<0.001**	**<0.001**	34.0	32.8	**<0.001**	**<0.001**	38.3	37.2	**<0.001**	**<0.001**
	diameter	0.83	0.80	0.376	0.444	0.83	0.81	0.372	1.000								
	cytotype:diameter	0.00	0.00	0.983	0.983												
	residuals	56.2				56.2				58.0				72.0			

[a] = Analysis includes data only from 58 plants for which diameter was measured.
[b] = Analysis includes data from 72 surveyed plants.
[c] = *P*-value corrected for 5% false discovery rate.

APPENDIX S5

Geographic distribution of species in the *Asphondylia auripila* group

Table S5.1. Summary of geographic distributions and *Larrea tridentata* cytotypes attacked by the species in the *Asphondylia auripila* group.

***Asphondylia* species**	**Reference**	**Organ attacked**	**Distribution**					**Host cytotypes attacked**		
			South Chihuahuan	**Central Chihuahuan**	**North Chihuahuan**	**Sonoran**	**Mojave**	**Diploid**	**Tetraploid**	**Hexaploid**
A. sp. "acuminata"	This study	bud	+	+	+			+		
A. apicata	Gagné & Waring (1990)	bud				+	+		+	+
A. auripila	Gagné & Waring (1990)	stem		+	+	+	+	+	+	+
A. barbata	Gagné & Waring (1990)	leaf (adaxial)		+	+	+	+	+	+	+
A. bullata	Gagné & Waring (1990)	bud			+	+	+	+	+	+
A. clavata	Gagné & Waring (1990)	leaf (adaxial)	+	+	+	+	+	+	+	+
A. digitata	Gagné & Waring (1990)	leaf (abaxial)			+	+	+	+	+	+
A. discalis	Gagné & Waring (1990)	leaf (abaxial)		+	+	+	+	+	+	+

A. fabalis	Gagné & Waring (1990)	leaf (abaxial)				+	+		+	+
A. florea	Gagné & Waring (1990)	bud				+	+		+	+
A. foliosa	Gagné & Waring (1990)	stem	+	+	+	+	+	+	+	+
A. sp. "hirsuta"	This study	leaf (abaxial)	+	+	+			+		
A. pila	Gagné & Waring (1990)	leaf (adaxial)				+	+		+	+
A. resinosa	Gagné & Waring (1990)	stem			+	+	+	+	+	+
A. rosetta	Gagné & Waring (1990)	bud				+	+		+	+
A. silicula (typical)	Gagné & Waring (1990)	leaf (abaxial)	+	+	+	+	+	+	+	+
"adaxial silicula"	This study	leaf (adaxial)								
A. villosa	Gagné & Waring (1990)	leaf (adaxial)				+	+		+	+

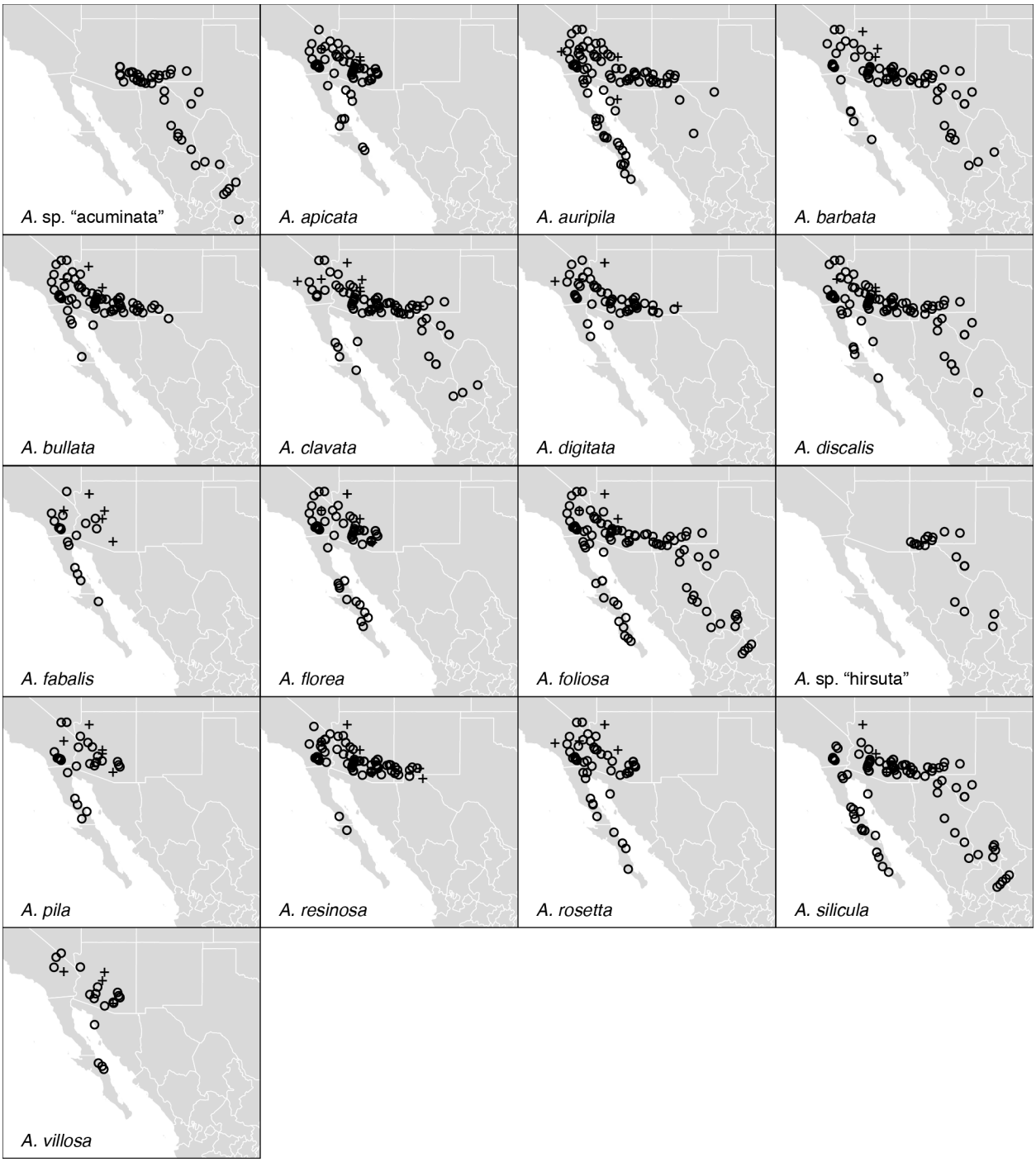

Figure S5.1. Geographic distribution of species in the *Asphondylia auripila* group. Collections from this study are plotted with open circles. Records from the literature (Werner & Olsen, 1973; Waring & Price, 1989; Gagné & Waring, 1990; Schowalter et al., 1999; Huggins, 2008) are plotted as “+”.

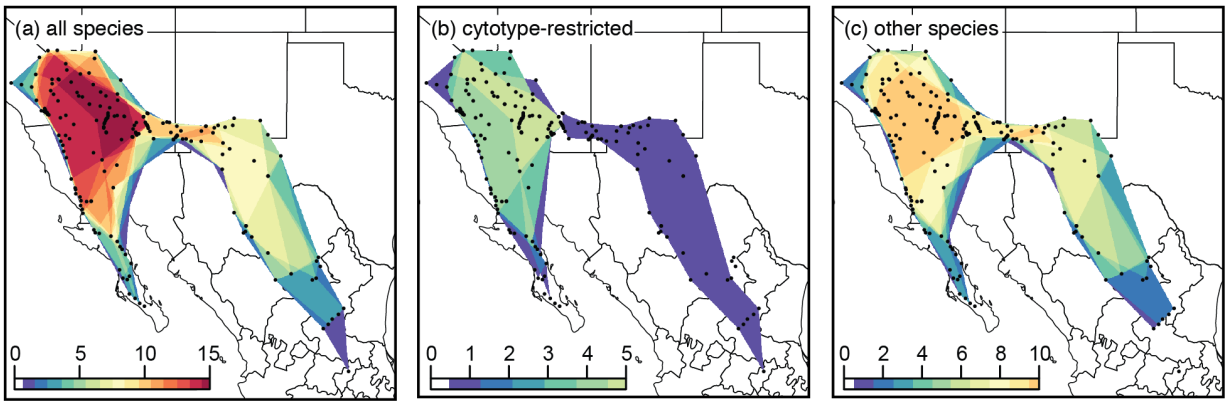

Figure S5.2. Regional species richness of *Asphondylia auripila* group. (a) All 17 species. (b) Cytotype-restricted species only. (c) All other *Asphondylia* species. Maps exclude the adaxial *A. silicula* gall morphotype, as its distribution is poorly known.

CHAPTER 2: Host plant and geography shape contrasting patterns of divergence in two specialized herbivores of creosote bush (*Larrea tridentata*)

Anticipated co-authorship: Kelsey M. Yule[1], Noah K. Whiteman[2]
1. Biodiversity Knowledge and Information Center, Arizona State University, Tempe, AZ 85282
2. Department of Integrative Biology, University of California Berkeley, Berkeley, CA 94706

ABSTRACT

Identifying factors that constrain migration – and how they act in geographic context – is essential to understand population divergence. The relative contributions of physical barriers vs. barriers mediated by species interactions is an open question in this rapidly advancing field. We investigated the effect of geography and host plant use on the population structure of two herbivores of creosote bush (*Larrea tridentata*, Zygophyllaceae). Creosote bush comprises three reproductively isolated and phenotypically differentiated cytotypes that are parapatrically distributed in the warm deserts of North America. In Chapter 1 we demonstrated that interactions with herbivores differ among cytotypes, which can limit herbivore distributions irrespective of physical or abiotic constraints. We therefore hypothesized that host plant use may also be associated with divergence between populations of widespread herbivore species, and as such, the well-defined contact zones between cytotypes may be discrete migration barriers analogous to mountain ranges. To address this hypothesis, we used RADcap sequencing to genotype hundreds of individuals from two herbivores (*Bootettix argentatus* and *Insara covilleae*) collected from across the range of each species. We then inferred phylogenetic networks and characterized population structure for each species to evaluate a correspondence between genetic divergence and host plant use. We next formally compared the overall effects of geography and host plant on genetic differentiation with distance-based redundancy analysis. To resolve the contributions of particular features to population divergence, we inferred the location of migration barriers using effective migration surfaces and identified concordance with physical barriers and cytotype contact zones. This allowed us to compare 1) the strength of biotic and physical barriers within species, and 2) the location and strength of barriers between species. We found that although geography and host plant predicted genetic divergence in both species, the degree and pattern of population structure differed markedly between them. Cytotype contact zones were among the strongest barriers to migration in *Bootettix*. These and many well-defined physical barriers resulted in a highly dissected pattern of genetic diversity across the species' range. By contrast, divergence across cytotype contact zones was weak or nonexistent in *Insara*, and genetic differentiation was low overall. Only a single major mountain corridor (the Madrean Archipelago) presented a significant barrier to migration. Our results demonstrate that biotic interactions can create substantial barriers to gene flow, rivaling or exceeding those of physical barriers, but that their effect may depend on the natural history of each species.

INTRODUCTION

Migration (gene flow) plays a central role in the generation and geographic distribution of biodiversity. When migration is local (i.e., occurring on a scale smaller than a species' range), isolation by distance (IBD) produces a gradient of genetic differentiation across space (Wright, 1943; Slatkin, 1993). Discrete barriers to migration can layer additional, localized genetic differentiation upon this background of IBD (i.e., population structure). Because drift and selection act upon subdivided populations independently, migration barriers permit neutral differentiation, local adaptation, and speciation (Ronce, 2014). Identifying factors that constrain migration – and how they act in geographic context – is thus essential to understand spatial variation in genetic diversity.

Although investigations of population structure have historically emphasized geographic factors that limit dispersal (e.g., distance or discrete barriers such as mountain ranges), the importance of isolation by environment (IBE) has recently come to the fore (Shafer & Wolf, 2013; Sexton et al., 2014; Wang & Bradburd, 2014). It has now been demonstrated in a range of organisms (Lowry et al., 2008; Gompert et al., 2014; Weber et al., 2016) that divergent natural selection upon populations in different environments or biased dispersal can promote or maintain population divergence irrespective of geographic distance. The way in which the environment limits migration may take several forms. Migration may continuously vary with "environmental distance" in a manner analogous to IBD (e.g., along abiotic gradients) or it may be disrupted by discrete environmental transitions (e.g., at the range limits of interacting species). The tendency of geography and the environment to covary has made disentangling their contributions to genetic divergence a longstanding challenge. Recently, new methodological approaches to partitioning the effect of confounded variables (Bradburd et al., 2013; Wang et al., 2013) and to identifying the location of migration barriers (Petkova et al., 2016) have invigorated this line of research.

One research direction concerns the predictability of IBE and the mechanisms that underlie it. Recent work on stickleback fish has shown that IBE due to biased dispersal consistently predominates over IBD in lake-stream population pairs (Weber et al., 2016). At broader evolutionary scales, comparisons across species have used new methods to identify both concordant and discordant patterns of divergence across physical barriers (Barratt et al., 2018), addressing classic questions in phylogeography (Avise, 1998) with a more expansive and precise toolkit. These and other recent examples have focused primarily on the effects of geography vs. the abiotic environment.

A parallel line of research has examined the role of species interactions in driving genetic divergence, especially those between herbivorous insects and their host plants (Feder et al., 1988; Via, 1999; Drès & Mallet, 2002; Nosil et al., 2002). This field has also benefitted from recent methodological advances, revealing a contribution of host plants but the primacy of geography in structuring many herbivore populations (Bakovic et al., 2019; Driscoe et al., 2019; Vidal et al., 2019). Both classic and recent work have typically considered the effect of host plants found in sympatry (e.g., Feder et al., 1988) or in a patchy mosaic (e.g., Nosil et al., 2002). However, host plants may also be parapatrically distributed with minimal geographic overlap. In such cases, the contact zone between host plants may be an ecological barrier to migration analogous to a mountain range. Establishing an effect of host plant in these systems requires careful consideration of migration near host plant contact zones, as well as statistical approaches that partition genetic

variance between confounded variables. To our knowledge, few studies have quantified the effect of geographic vs. biotic migration barriers and compared their effects across species.

Herbivorous insects are excellent systems in which to compare the effect of geography and environment on population divergence. Genetic differentiation between populations on different host plants is widespread in herbivorous insects (reviewed in Forbes et al., 2017). Furthermore, communities of insects on shared host plants provide an opportunity to compare how a common set of geographic and environmental conditions affect genetic differentiation in different taxa. Do communities diverge as a unit, as has been found in arthropods associated with pitcher plants (Satler & Carstens, 2016, 2017; Satler et al., 2016)? Or are responses instead idiosyncratic and related to differences in life history (Phillipsen et al., 2015)?

To address these questions, we investigated the effect of geography and environment on the population structure of two herbivores of creosote bush (*Larrea tridentata*, Zygophyllaceae), a dominant shrub in the warm deserts of North America. The creosote bush grasshopper (*Bootettix argentatus*, Orthoptera: Acrididae) and creosote bush katydid (*Insara covilleae*, Orthoptera: Tettigoniidae) consume only creosote bush (Rhoades & Cates, 1976; Chapman et al., 1988) in the Chihuahuan, Sonoran, Mojave, and Peninsular Deserts (Fig 1A-B). Across this range, the nominal species *L. tridentata* comprises three reproductively isolated cytotypes (populations of different ploidal level) that arose through autopolyploidy (whole genome duplication without hybridization). As is often observed in autopolyploid complexes (Lewis, 1980; Levin, 1983; Ramsey & Schemske, 2002), the diploid, tetraploid, and hexaploid cytotypes of creosote bush differ phenotypically (Yang, 1967; Barbour, 1968; Bohnstedt & Mabry, 1979; Laport & Ramsey, 2015; Laport et al., 2016) and are geographically segregated (Barbour, 1969; Yang, 1970) (Fig 1C). We previously found that many members of a specialized gall midge community discriminate among creosote bush cytotypes, and cytotype contact zones constrain herbivore distributions (O'Connor et al., 2019). Together, these observations suggest the hypothesis that contact zones between cytotypes are biotic migration barriers.

A long history of comparative biogeography in North American Deserts has also identified many physical features that may be contemporary or historical barriers to migration (reviewed in (Hafner & Riddle, 2011), Fig. 1C). While herbivore populations are currently contiguous across most of these boundaries, Late-Pleistocene glaciation or interglacial pluvials (wet periods) likely fragmented and displaced populations near cooling mountain ranges and expanding lacustrine

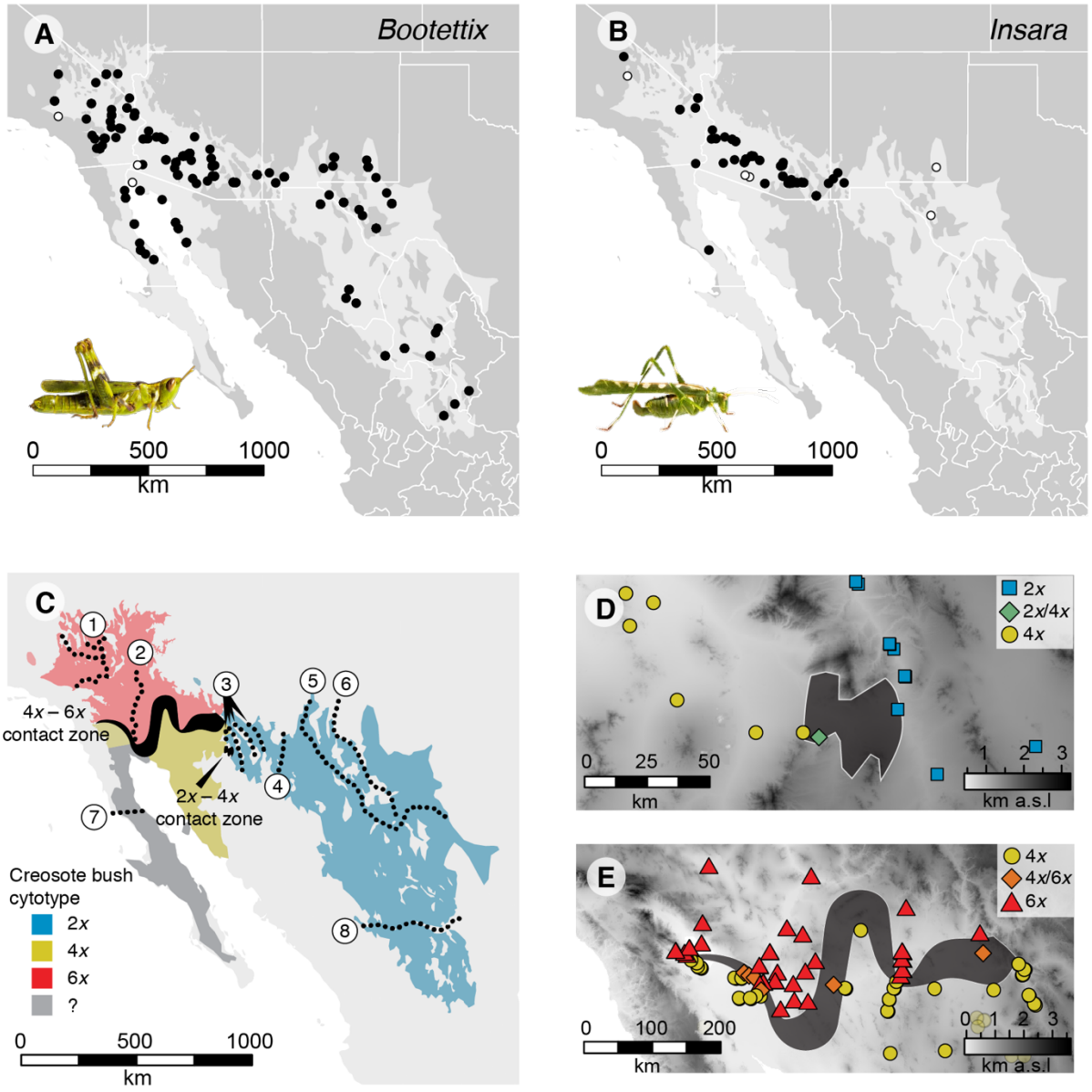

Figure 1. Sample localities and potential migration barriers. **A.** Location of 425 *Bootettix argentatus* samples collected from 116 sites. Open points indicate collections that were not included in the final dataset (see text for details). Light gray indicates geographic distribution of creosote bush. **B.** Location of 116 *Insara covilleae* samples collected from 53 sites. **C.** Distribution of creosote bush cytotypes and location of hypothesized barriers to gene flow. Host plant contact zones are indicated in black polygons, where polygon width roughly indicates uncertainty in location. Physical barriers are shown with dotted lines and numbered. 1: Mojave River Lacustrine system. 2: Colorado River. 3. Madrean Archipelago. 4: Deming Plains. 5: Rio Grande. 6: Sacramento / Guadalupe / Cathedral / Davis / Chisos Mountains. 7: Mesa Huatamote. 8: Southern Coahuila Filter Barrier. **D.** Inset of contact zone between diploid and tetraploid creosote bush indicated in panel C. Points represent cytotype of local creosote bush populations. km a.s.l. = kilometers above sea level. **E**. Inset of contact zone between tetraploid and hexaploid creosote bush indicated in panel C. Cytotype data from (Laport et al., 2012; Laport & Minckley, 2013; Laport & Ramsey, 2015).

systems (Hunter et al., 2001; Wilson & Pitts, 2010; Hafner & Riddle, 2011). In addition to isolation by distance, we hypothesized that these geographic barriers would shape the geographic structure of creosote bush herbivores.

To evaluate the contribution of geographic and biotic factors to population divergence, we genotyped range-wide collections of *Bootettix* and *Insara* using RADcap sequencing (Hoffberg et al., 2016). With the resulting SNP data, we first inferred phylogenetic networks for each herbivore species and mapped on host plant use to evaluate long-term associations between plant cytotypes and herbivore lineages. We then used distance-based redundancy analysis to partition the relative effect of geographic distance and host plant on genetic distance among individuals. Next, we visualized genetic gradients and discontinuities within each species with principal components analysis. Finally, we inferred spatial variation in migration rates and compared regions of reduced migration to hypothesized geographic and biotic migration barriers. Our results revealed a role for both geography and host plant in population structure of both herbivores, but the strength and pattern of subdivision differed markedly between species.

METHODS

Collections

We collected samples from across the range of *Bootettix argentatus* (Fig 1A, 425 individuals from 116 sites, max = 15, min = 1, median = 2) and *Insara covilleae* (Fig 1B, 116 individuals from 53 sites, min = 1, median = 1) between July 2015 and September 2018. Our sample selection strategy aimed to maximize geographic breadth and density, which aids in detecting IBE (Wang & Bradburd, 2014) and localizing genetic discontinuities among populations. We placed particular emphasis on known contact zones between creosote bush cytotypes. We complemented this breadth with deeper sampling at a subset of sites (*Bootettix:* 35 sites with ≥ 5 individuals, 17 sites with ≥ 10 individuals; *Insara*: six sites with ≥ 5 individuals, two sites with ≥ 10 individuals), allowing calculation of population-level genetic divergence. We also collected four congeners of *I. covilleae* (*I. elegans*) from Arizona, New Mexico, and Texas. Samples were stored in 100% ethanol and maintained at -20˚C until further processing.

Scoring host-plant use

We did not directly determine creosote bush cytotype at our sampling locations. However, flow cytometry of creosote bush populations from across the species' range (Laport et al., 2012; Laport & Minckley, 2013; Laport & Ramsey, 2015) has identified broad regions occupied by a single cytotype and delimited the boundaries between them. Diploid and tetraploid cytotypes meet in a short and well-defined contact zone in southeastern Arizona (Fig. 1D), while tetraploids and hexaploids meet in a longer contact zone that extends from central Arizona to southern California (Fig. 1E) (Laport et al., 2012). This previous work allowed us to indirectly assign host plants to most herbivore populations based on location. We assigned a host plant to an herbivore if its sampling locality was separated from a cytotype contact zone by at least one single-ploidy plant population. If not, we scored the host plant as ambiguous (e.g., "4*x* / 6*x*"). Because creosote bush populations in Baja California and the Colorado River delta have not been surveyed with flow cytometry, we omitted these samples from dbRDA analyses (see below).

DNA extraction

We extracted DNA from all samples using single-tube or 96-well DNeasy Blood and Tissue Kits (Qiagen). For adults and late-instar juveniles, we rehydrated one femur in autoclaved ddH20, chilled it to -80˚C, then homogenized for 1-4 minutes at 25-50 Hz. For early instar juveniles we homogenized whole bodies. We followed manufacturer's protocols with the following exceptions: samples were digested with proteinase K overnight and treated with RNAse A prior to binding to silica columns.

Pilot sequencing

To determine the suitability of standard double-digest RAD (ddRAD) sequencing for our project, we sequenced 49 *Bootettix* and 17 *Insara* collected across the geographic breadth of each species. Sequencing libraries were prepared as in Peterson et al., (2012) with slight modifications (see Barker et al., 2017). We selected two 6-cutter enzymes (EcoRI and PstI) to minimize the number of RAD loci produced. Orthopteran genomes are typically large (5-20 Gb, Gregory, 2019), so digestion with more frequent cutters would result in an excess of sequenced loci and reduced coverage for a given sequencing effort (Peterson et al., 2012). We further attempted to limit the number of sequenced loci by selecting a narrow size range of digested fragments for sequencing (90 bp). *Bootettix* and *Insara* samples were sequenced on a single HiSeq 4000 lane (PE 100), with each sample allocated 1/96 of a lane. We assembled RAD loci for each species using ipyrad v0.7.19 (Eaton, 2014), clustering loci at 90% identity and retaining loci found in ≥ 2 individuals.. We used default values for all other assembly parameters.

Despite efforts to minimize the number of sequenced loci, we assembled 602,993 loci for *Bootettix* and 180,0128 loci for *Insara*. Mean sequencing depth per locus was typically acceptable within individuals (*Bootettix* = 7×, *Insara* = 13×), but locus coverage across samples was typically low. The high degree of missing data indicated that the sequencing effort required for a large-scale population genomics study with ddRAD would be cost-prohibitive.

RADcap sequencing

Bait design

In order to economically genotype hundreds of samples, we instead performed RADcap sequencing (Hoffberg et al., 2016). RADcap combines 3RAD – a variant of RAD sequencing (Bayona-Vásquez et al., 2019) – with targeted sequence capture to enable reliable, efficient, and high-throughput sequencing of RAD loci. This approach has several advantages, the first of which comes during initial library preparation. Like ddRAD sequencing, 3RAD uses two primary restriction enzymes to digest sample DNA. Unlike ddRAD, however, 3RAD uses a third restriction enzyme to cleave adapter dimers during the ligation stage, increasing the efficiency of sample-to-adapter ligation. This approach further increases efficiency and minimizes sample loss by performing digestion and ligation in a single tube. Combinatorial indexing with two in-line barcodes and one indexing read permits massive multiplexing, while an 8 bp unique molecular index facilitates the identification and removal of PCR duplicates.

The second advantage of RADcap comes with sequence capture. By using biotinylated RNA baits designed to bind to known RAD loci, target loci can be physically separated from the remainder of the genome and sequenced. This increases the repeatability of genotyping across individuals

and minimizes the sequencing required for sufficient coverage (Ali et al., 2016; Hoffberg et al., 2016).

We therefore designed sequence capture baits from pilot sequencing data with a custom pipeline coded in Perl. We began by identifying suitable targets from among the full set of RAD loci assembled for each species, retaining loci that met the following criteria:

1) The locus was sequenced at least once in samples of divergent populations east and west of the Madrean Archipelago (to minimize allelic dropout).
2) The locus lacked an NsiI recognition site, as this restriction enzyme was used to cleave adapter dimers during 3RAD library preparations. Loci with the NsiI recognition site would be digested during library preparation and unsequenceable.
3) No sequenced alleles were > 5% divergent from the consensus sequence of that locus, as capture efficiency declines precipitously above 5% divergence (Bi et al., 2012).
4) The locus contained no indels, which can impede bait binding (Arbor Biosciences, 2017).
5) The locus consensus sequence had GC content between 30% and 70%, as extreme GC content can result in poor sequence capture efficiency (Bi et al., 2012).

We then performed an all vs. all BLAST search of retained loci for each species and removed any locus with a significant hit using a liberal significance threshold (e-value ≤ 10). We further removed loci containing TE or simple repeat motifs identified with RepeatMasker v4.0.7 (Smit et al., 2015) and the RepBase arthropod repeat library (Bao et al., 2015). These steps minimized the possibility that baits would cross-bind or capture non-target loci with similar repeat motifs.

We used the same pipeline to design baits for two additional species for separate projects. As in Hoffberg et al. (2016), we intended to combine bait sets for multiple species into a single custom bait order to minimize project cost. We therefore performed an additional all vs. all BLAST search of loci from all four species and removed those with significant hits. Finally, we randomly selected 10,000 loci from each species and designed a single 80-bp bait sequence that was complementary to each target locus.

We ordered 40,000 myBaits sequence capture baits from Arbor Biosciences (Ann Arbor, Michigan). Baits for each species were therefore at 0.25× concentration relative to the manufacturer's recommendation. Using an approach similar to RADcap (Rapture), Ali et al. (2016) found that capturing RAD loci with 0.2× bait concentrations can yield sequencing performance that is indistinguishable from libraries captured with 1× baits.

Library preparation

We prepared sequencing libraries following 3RAD protocols developed by the Glenn lab at UGA (http://baddna.uga.edu) and custom adapters designed to match our set of restriction enzymes. Prior to digestion, we quantified DNA concentration of each sample with Quant-iT PicoGreen dsDNA kit (Invitrogen) and a BioTek Synergy microplate reader. After diluting all samples to 10 ng / µL, we set up an initial digestion for all samples. Digestions consisted of: 10 µL DNA, 0.5 µL PstI-HF (NEB), 0.5 µL EcoRI-HF (NEB), 0.5 µL NsiI-HF (NEB), and 1.5 µL 10× CutSmart buffer (NEB). We then added a unique pair of barcoded adaptors to each sample defined its position in the 96-well plate. Eight unique "NsiI" adaptors (which ligate to DNA cut by PstI but are cleaved

by NsiI if self-ligated) corresponded to the plate's eight rows, while twelve "EcoRI" adapters (which ligate to DNA cut by EcoRI) corresponded to the twelve columns as detailed in (Bayona-Vásquez et al., 2019), We incubated this mixture at 37˚C for 1 hour, then added the following ligation mixture to each sample: 2.75 µL molecular grade water, 1.5 µL ATP (NEB), 0.5 µL 10× ligase buffer (NEB), and 0.25 µL DNA ligase 400 u/mL (NEB). The combined digestion / ligation mixture was then cycled twice between ligation and digestion conditions (22˚C for 20 min, 37˚C for 10 min, 22˚C for 20 min, 37˚C for 10 min) before heat-killing the enzymes for 20 min at 80˚C.

We retained 10 µL from each reaction as a backup, then pooled the remaining 10 µL from each sample within a plate. Each plate was pooled separately. Out of necessity, some of our digest / ligation plates contained samples for multiple species. When this occurred, we pooled each species separately. We cleaned pooled reactions with 1.25× SeraMag beads (prepared as in Rohland and Reich 2012) and resuspended samples in a total volume of 60µL molecular grade water.

For each pool, we next performed 6 replicate single-cycle PCR with an "iTru5 8N" primer containing a random 8 bp sequence and partial Illumina P5 adaptor (see Bayona-Vásquez et al., 2019). Each reaction contained: 25 µL KAPA HiFi HotStart ReadyMix (Roche), 10 µL pooled and cleaned DNA, and 5 µL iTru5 8N primer at 5µM. The single reaction cycle was as follows: 98˚C for 1 min, 60˚C for 30 sec, 72˚C for 6 min. We pooled replicate reactions and cleaned with 1.5× SeraMag beads, resuspending in 33 µL molecular grade water.

We next completed the Illumina adaptors and amplified libraries with triplicate reactions for each pool of DNA. Reactions consisted of: 25 µL KAPA HiFi HotStart ReadyMix (Roche), 5.0 µL P5 primer, 5.0 µL "iTru7" primer with a 8 bp plate-level barcode, and 10 µL DNA resulting from pooled and cleaned single-cycle PCRs. Reaction conditions were: 98˚C for 2 min, followed by 12 cycles of 98˚C for 20 sec, 60˚C for 15 sec, 72˚C for 30 sec, and a final 72˚C elongation for 5 min. We pooled and cleaned reactions with 1.5× SeraMag beads, resuspending in 60 µL.

We quantified resulting DNA libraries with the Qubit HS dsDNA kit (Invitrogen) and calibrated the quantity of DNA to use in sequence capture. We expected the ratio of target to non-target DNA in our samples to be low, since our baits targeted a minority of RAD loci generated by double-digestion. In addition, the reduced relative concentration of baits further limited the opportunities for bait binding. In such conditions, increasing total amount of DNA in capture reactions can improve capture performance (McCartney-Melstad et al. 2016, Arbor Biosciences 2018). We therefore used between 1.5 and 2 µg of DNA per capture reaction.

We performed sequence capture following manufacturer's instructions (myBaits protocol v4, Arbor Biosciences) with two deviations. First, we used 10 µL Roche SeqCap EZ Developer Reagent (Roche) in lieu of Block C and Block O provided with the myBaits kit. For invertebrates, the developer reagent may provide better blocking of repetitive DNA (Ke Bi, personal communication). DNA mixed with 10 µL Developer Reagent and 0.5 µL Block A (provided with the myBaits kit) were dried to a volume of 7.5 µL in a SpeedVac prior to mixing with the hybridization mixture (step 1.4 in myBaits protocol v4). Second, we hybridized baits to DNA for 52 hours at 65˚C.

We directly amplified bead-bound libraries following capture. Triplicate reactions consisted of 25 µL KAPA HiFi HotStart ReadyMix (Roche), 5.0 µL P5 primer, 5.0 µL "iTru7" primer with 8 bp plate-level barcode, and 10 µL DNA. Reaction conditions were: 98˚C for 2 min, followed by 16 cycles of 98˚C for 20 sec, 60˚C for 15 sec, 72˚C for 30 sec, and 72˚C elongation for 5 min.

All samples for this project and several additional projects were sequenced across portions of two Illumina HiSeq 4000 lanes (PE100) at the Vincent J. Coates Sequencing Laboratory at the University of California, Berkeley. In total, we pooled 480 or 609 individuals per lane at equimolar concentrations.

RAD assemblies

We used ipyrad v0.9.13 (Eaton, 2014) to quality-filter and assemble raw sequencing reads into de novo RAD loci using default parameters except as follows: min length of quality-trimmed reads = 80, clustering threshold = 0.9, max proportion of variable sites = 0.2, max shared heterozygotes per locus = 0.25, min number of samples per locus = 4.

To create datasets for population structure analyses, we chose a single biallelic SNP per RAD locus with minor allele frequency ≥ 2.5%. This frequency threshold balanced biases introduced by overly-stringent minor allele filters and the inclusion of singletons (Linck & Battey, 2019). Prior to population structure analyses, we filtered loci violating HWE in ≥ 2 sampled populations. We then iteratively filtered samples and loci to produce datasets in which all samples were sequenced at ≥ 75% of loci, and all loci were sequenced in ≥ 90% of individuals. Filtering was performed with custom Perl and R code, as well functions from the 'adegenet' package v2.1.1 (Jombart, 2008) in R v3.6.1 (R Core Team, 2019).

Phylogenetic network inference

We expected that a potentially complex history of divergence and gene flow between herbivore populations would violate the assumption of a strictly bifurcating phylogeny. To better represent such reticulation, we inferred phylogenetic networks for *Bootettix* and *Insara* individuals using the neighbor-net algorithm implemented in SplitsTree v4.15.1 (Huson & Bryant, 2006).

Principal Components Analysis (PCA)

We summarized genetic variation among samples using principal components analysis (PCA) implemented in 'adegenet'. Missing genotypes were imputed with mean allele frequencies. To visualize genetic variation across the range of each species, we transformed PCA coordinates into color space coordinates with the 'colorplot' command and mapped resulting colors upon sample localities.

Calculating population-level differentiation

We summarized population-level divergence with the Weir and Cockerham estimator of F_{ST} (Weir & Cockerham, 1984) implemented in adegenet. We calculated F_{ST} only between populations with ≥ 10 individuals.

Estimating effective migration surfaces

We visualized spatial variation in migration rates for each species using EEMS (Petkova et al., 2016). EEMS models gene flow between demes as a function of geographic distance and a

continuously-varying effective migration surface. When plotted upon a map, effective migration surfaces provide an intuitive way to visualize barriers to gene flow across a landscape and compare such barriers to geographical and ecological features. After exploring EEMS models with 100-1,000 demes, we performed final analyses with 1,000 demes. We ran three independent chains of each model for 4,000,000 generations with a 1,500,000 generation burn-in, sampling every 1,000th generation. Following the advice of (Petkova et al., 2016), we tuned EEMS hyperpriors to achieve MCMC acceptance rates in the range of 15-30% for each model parameter.

Isolating the effects of host plant and geography

We next quantified the contributions of host plant use and geographic distance to genetic distance with distance-based redundancy analysis (dbRDA, McArdle & Anderson, 2001) using the 'vegan' v.2.5-6 (Oksanen et al., 2017) and 'adespatial' v.0.3-7 (Dray et al., 2019) packages in R. dbRDA quantifies the relationship between a distance matrix (pairwise genetic distance) and linear predictor variables (host plant and geographic location). Within this framework, we quantified the effect of each predictor separately (Y ~ X1; Y ~ X2), both predictors together, (Y ~ X1+X2), as well as the marginal effect of each predictor after conditioning upon the other (Y ~ X1 | X2). The joint effect of geography and host plant is the variance explained by a model including both predictors but cannot be attributed to the marginal effect of either (Fig. 2).

We calculated pairwise genetic distances among individuals as the mean deviation in allele frequency across all loci. Missing genotypes were imputed with mean allele frequencies to ensure a Euclidean distance matrix (implemented with the "bed2diffs_v2" function provided with EEMS (Petkova et al., 2016)).

We summarized the geographic arrangement of localities with distance-based Moran eigenvector maps (dbMEMs, Borcard & Legendre, 2002). This approach uses principal coordinates analysis of a truncated geographic distance matrix to capture the localized neighbor relationships between samples. The resulting set of linear variables provide a more effective summary of geography for linear modeling than raw latitude and longitude. We calculated dbMEMs using the "dbmem" command in 'adespatial'. Following best practices outlined by Borcard et al. (1992), we truncated the distance matrix at the maximum nearest-neighbor distance, replacing all larger values with four times the maximum nearest-neighbor distance. We retained dbMEMs that 1) had positive eigenvalues, and 2) had a significant association ($P < 0.05$) with genetic dissimilarity (as in Driscoe et al., 2019). We used a matrix of retained dbMEMs to represent geography in final dbRDA analyses.

Because not all samples could be confidently assigned a host plant cytotype, we tested the effect of this uncertainty on our inference by conducting with and without the ambiguously coded samples. Results were quantitatively similar, so we present only results that include ambiguous cytotypes for simplicity.

We used the "varpart" function in 'vegan' to perform variance partitioning and evaluated the significance of each partition using dbRDA and permutational ANOVA (9,999 free permutations).

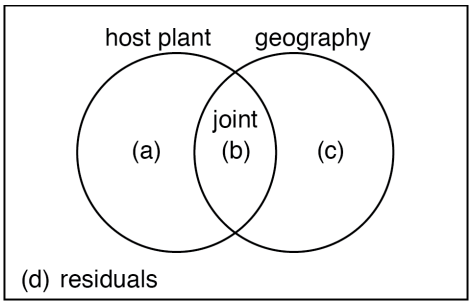

Figure 2. Schematic illustration of variance partitions. Partition letters are referenced in in Tables 1-2.

RESULTS

Sequencing performance

Two lanes of Illumina HiSeq 4000 PE100 sequencing resulted in 361M and 247M reads passing filter. We obtained a median of 360,005 reads across 425 *Bootettix* individuals (range = 26,513 – 1,788,619; s.d. = 329,968) and 469,114 reads per 116 *Insara* individuals (range = 34,334 – 2,432,770; s.d. = 371,350).

RAD assembly and SNP filtering

Prior to additional filtering for sample or locus coverage, our de-novo assemblies included 86,211 RAD loci in *Bootettix* and 26,461 loci in *Insara*. The disparity between the number of targeted loci (10,000 per species) and number of assembled loci is likely due to a combination of 1) off-target bait binding and 2) incomplete removal of unbound RAD loci during RADcap library preparation. Individual *Bootettix* samples were genotyped at a median of 12,201 loci (range = 620 – 37,101; s.d. = 7,242) to a median depth of 22.6×, and *Insara* samples were genotyped at a median of 12,683 loci (range = 608 – 30,379; s.d. = 4,803) to a median depth of 29.7×.

After filtering for minor allele frequency, violations of HWE, and selecting a single SNP per locus, our *Bootettix* dataset included 351 individuals and 732 SNPs and the *Insara* assembly included 108 individuals and 1,513 SNPs.

Despite our attempt to capture loci shared between divergent populations of *Bootettix*, we noticed that many pre-filtered RAD loci were genotyped exclusively in eastern or western populations. We hypothesize that this allelic dropout is due to restriction-site mutations unique to divergent *Bootettix* lineages. To increase the size of our SNP datasets, we therefore re-assembled and filtered RAD loci for eastern and western *Bootettix* populations separately. Our final SNP set for western *Bootettix* included 202 individuals and 1,879 loci while the eastern dataset included 152 individuals and 1,685 loci.

During exploratory analyses we noticed several unexpected patterns of genetic variation in *Insara*. First, three *Insara covilleae* samples from the eastern margin of the species' range clustered with *I. elegans* (a well-differentiated congener) in a phylogeny. These samples appeared to have been misidentified and were removed from further analysis. Second, heterozygosity differed systematically between male and female *Insara* at a subset of loci. Such a pattern is expected for loci located on sex chromosomes (males = X0, females = XX). Because we wished to include males and females in a single set of population genomic analyses, we identified and excluded putatively sex-linked loci. To do so, we first performed a PCA of heterozygosity at each locus

(homozygote = 0, heterozygote = 1). As expected, this PCA segregated males and females along a single major axis of variation. We then identified the loci that loaded most heavily on this axis and removed them from subsequent analyses. A PCA of the remaining 986 SNPs showed no such sex bias in heterozygosity.

Phylogenetic clustering of host plant use

Host use was strongly clustered on the *Bootettix* phylogenetic network, with individuals collected from each creosote bush cytotype forming nearly exclusive lineages (Fig. 3). The greatest divergence was between diploid-associated samples in the Chihuahuan Desert and populations associated with polyploid creosote bush in the western deserts (Sonoran, Mojave, and Peninsular). Phylogenetic clustering was also strong among western populations. Individuals collected near cytotype contact zones fell discretely into "tetraploid-associated" and "hexaploid-associated" clades (Fig 4), consistent with the prediction that host plant cytotype structures host plant populations.

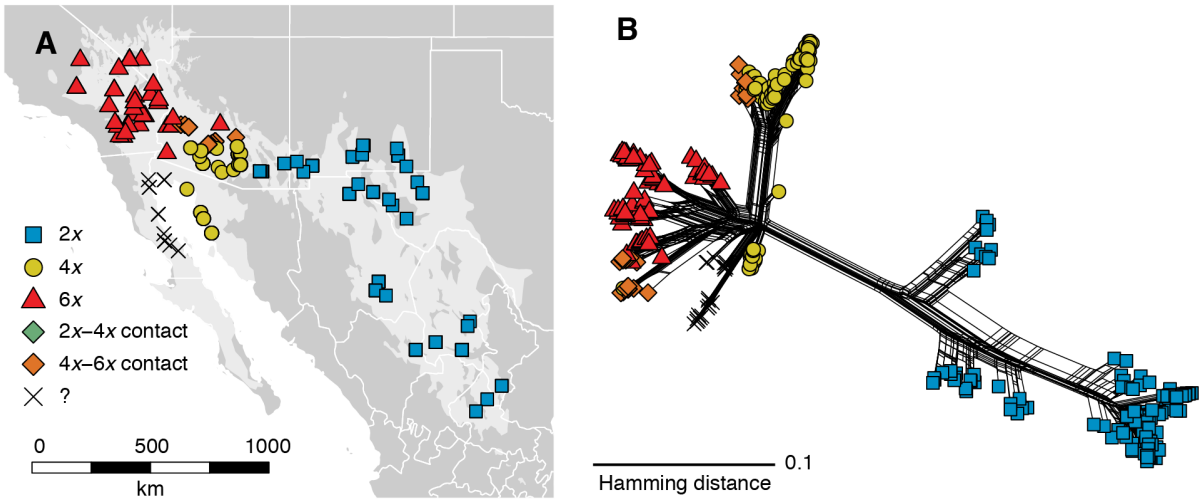

Figure 3. Host plant use is clustered on the *Bootettix* phylogenetic network. **A**. *Bootettix* sample locations and host plant assignments. **B.** Phylogenetic network demonstrating near-complete association of *Bootettix* lineages and host plant cytotypes.

By contrast, *Insara* populations showed much less phylogenetic structure – most branch length was concentrated at network tips, and internal branches were poorly resolved. Although there was some evidence for phylogenetic clustering of diploid hosts (Fig. 5), this association of host and lineage was imperfect. Individuals from higher-ploidy host plants showed no discernible association between host plant and phylogeny.

Effect of host plant and geography differ between herbivores

dbRDA analyses found that the contributions of host plant and geography to genetic variation differed between *Bootettix* and *Insara*. Host plant and geography were each strong predictors of genetic structure in *Bootettix*, and a model including both factors explained nearly 90% of genetic variance (Table 1, $R^2 = 0.892$, $P = 0.0001$). Roughly equal proportions of total genetic variance was explained by the marginal effects of host plant ($R^2 = 0.095$, $P = 0.0001$) and geography ($R^2 = 0.1$, $P = 0.0001$), but most genetic variance could not be uniquely ascribed to either variable (joint effect of host plant and geography $R^2 = 0.698$).

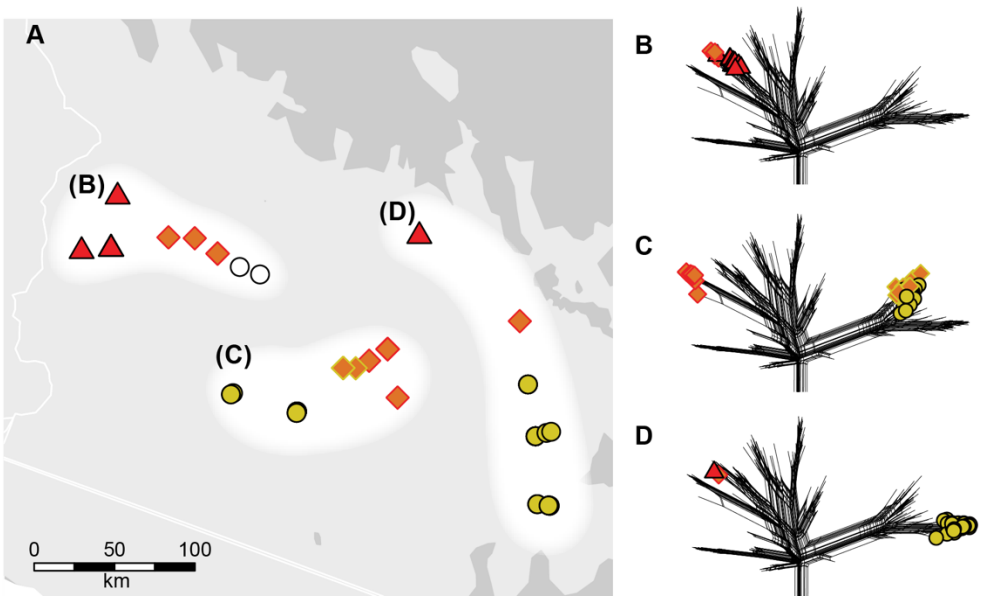

Figure 4. *Bootettix* collected near the tetraploid--hexaploid contact zone in Arizona come from distinct "tetraploid-associated" and "hexaploid-associated" lineages. **A.** Location of three transects near the diploid-tetraploid contact zone. Phylogenetic placement of samples in these zones are shown in panels B-D. Plotting characters indicate host plant assignments (as in Fig. 2). Samples with ambiguous host plants are outlined to indicate with which *Bootettix* lineage they cluster: yellow = "tetraploid associated," red = "hexaploid-associated". **B.** Populations from within the tetraploid–hexaploid contact zone in northwestern Arizona all cluster with hexaploid-associated *Bootettix*. Population density was high at these sites, but we were unable to locate *Bootettix* at two nearby sites within the known range of tetraploid creosote (open circles in panel A). **C.** There was an abrupt transition between tetraploid-associated and hexaploid-associated lineages within 10 km in central Arizona. **D**. Samples from a site with ambiguous host plants clustered with hexaploid-associated *Bootettix* in the eastern limit of the Sonoran Desert.

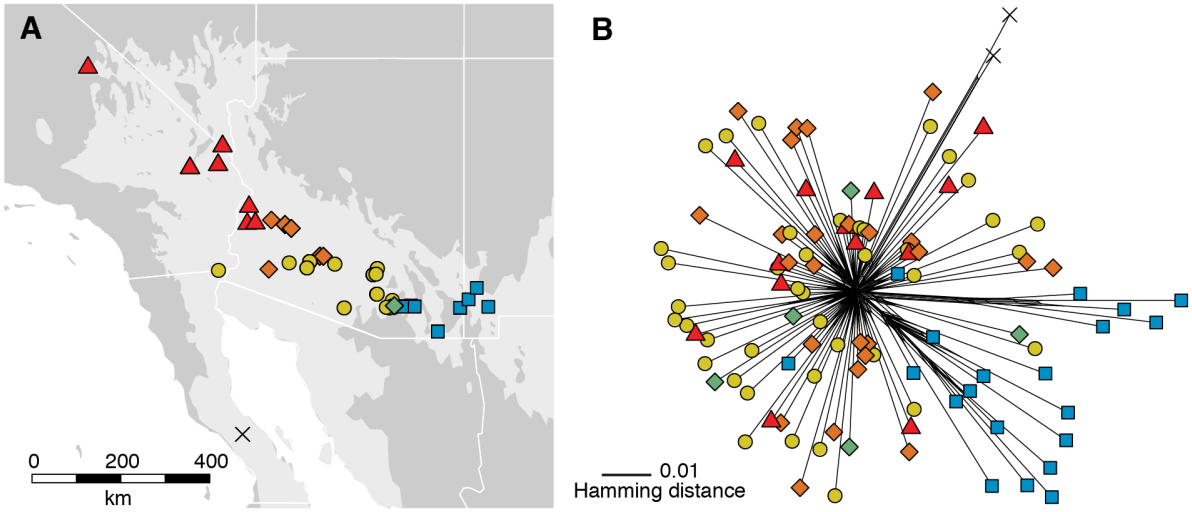

Figure 5. Host plant use shows limited clustering on the *Insara* phylogenetic network **A.** *Insara* sample locations and host-plant assignments. **B.** *Insara* from diploid creosote bush tend to cluster into two distinct lineages, while there is no discernible clustering of samples collected from tetraploid or hexaploid host plants. Plotting characters as in Fig. 3.

We next considered only western *Bootettix* populations, where geographic distance and host plant use is less confounded. Host plant and geography explained most genetic variance in western *Bootettix* ($R^2 = 0.830$, $P = 0.001$). Relative to the full dataset, and a larger proportion was attributed to the marginal effect of host plant ($R^2 = 0.154$, $P = 0.0001$) and geography ($R^2 = 0.279$, $P = 0.001$), but the majority of variance was again explained by their joint effect ($R^2 = 0.351$).

Host plant and geography explained much less genetic variance in *Insara* than in *Bootettix* ($R^2 = 0.101$, $P = 0.0001$). Most of this variance was due to the marginal effects of host plant ($R^2 = 0.025$, $P = 0.0001$) and geography ($R^2 = 0.055$, $P = 0.0001$), indicating that these variables are less confounded than in *Bootettix*. We next considered *Insara* collected only from tetraploid or hexaploid plants to test whether differences between higher-ploidy hosts contribute to population structure. Our model explained even less genetic variance in western individuals ($R^2 = 0.019$, $P = 0.0001$) than in the full dataset. The marginal effects of host plant ($R^2 = 0.005$, $P = 0.0415$) and geography ($R^2 = 0.013$, $P = 0.0001$) were significant, but had little predictive value.

Evidence of restricted migration at host plant contact zones: *Bootettix*

While dbRDA results effectively summarize the contribution of geography and host plant to genetic structure across the full species' range, they do not allow us to quantify and compare the effect of particular migration barriers. We therefore combined the results of PCA and EEMS to further explore migration near cytotype contact zones.

Consistent with phylogenetic network analysis, PCA revealed substantial population structure in *Bootettix*. The first principal component (explaining 46.9% of variance) separated eastern and western individuals (Fig. 6). Genetic differentiation was high between eastern and western populations overall (F_{ST}: 0.61 – 0.90), including between adjacent populations in Arizona (F_{ST}= 0.80, distance = 80 km, Fig. 6C).

The phylogeographic break between eastern and western *Bootettix* coincided with the 2*x* – 4*x* contact zone as well as the western edge of the Madrean Archipelago. We attempted to collect *Bootettix* closer to the contact zone in order to isolate the effect of host plant. However, across three years of fieldwork we did not encounter *Bootettix* at any of the 22 sites we visited. Nor did we find *Bootettix* at 11 sites near another 2*x* – 4*x* contact zone along the San Pedro River (Fig 6C).

We also found evidence that migration among western *Bootettix* populations is restricted at the 4*x* – 6*x* contact zone (Fig. 7). EEMS inferred a network of migration barriers within the Sonoran and Mojave Deserts (Fig. 7B), many of which corresponded with hypothesized barriers to gene flow. Notably, there was a close correspondence between migration barriers in southern California and Arizona and the 4*x* – 6*x* contact zone (Fig. 7C). EEMS inferred that effective migration rates across this barrier were $< 10^{-2}$ relative to the range-wide mean, among the lowest in the western deserts (Fig. 8). This is especially remarkable because the barrier does not coincide with any geographic features that might impede dispersal.

Table 1. Genetic variance in *Bootettix* partitioned among explanatory variables with dbRDA. The convention X1 | X2 indicates the effect of X1 after conditioning upon X2. Partitions are visually displayed in Fig. 2. Marginal effects of host plant and geography are given by partitions a and c respectively. Their joint effect (variance that cannot be uniquely ascribed to either factor) is given in partition b. The significance of the joint effect (partition b) is not testable with dbRDA.

dataset	factor(s)	partition	d.f.	Adjusted R^2	*P*
Bootettix all	host plant	a+b	3	0.793	**0.0001**
	geography	b+c	10	0.798	**0.0001**
	host plant + geography	a+b+c	13	0.892	**0.0001**
	host plant | geography	a	3	0.095	**0.0001**
	joint (host plant, geography)	b	0	0.698	-
	geography | host plant	c	10	0.1	**0.0001**
	residuals	d		0.108	-
Bootettix west	host plant	a+b	2	0.505	**0.0001**
	geography	b+c	8	0.63	**0.0001**
	host plant + geography	a+b+c	10	0.783	**0.0001**
	host plant | geography	a	2	0.154	**0.0001**
	joint (host plant, geography)	b	0	0.351	-
	geography | host plant	c	8	0.279	**0.0001**
	residuals	d		0.217	-

Table 2. Genetic variance in *Insara* partitioned among explanatory variables with dbRDA.

dataset	factor(s)	partition	d.f.	Adjusted R^2	*P*
Insara all	host plant	a+b	4	0.046	**0.0001**
	geography	b+c	4	0.077	**0.0001**
	host plant + geography	a+b+c	8	0.101	**0.0001**
	host plant | geography	a	4	0.025	**0.0001**
	joint (host plant, geography)	b	0	0.021	-
	geography | host plant	c	4	0.055	**0.0001**
	residuals	d		0.899	-
Insara west	host plant	a+b	2	0.006	**0.0048**
	geography	b+c	2	0.014	**0.0001**
	host plant + geography	a+b+c	4	0.019	**0.0001**
	host plant | geography	a	2	0.005	**0.0415**
	joint (host plant, geography)	b	0	0.001	
	geography | host plant	c	2	0.013	**0.0001**
	residuals	d		0.981	

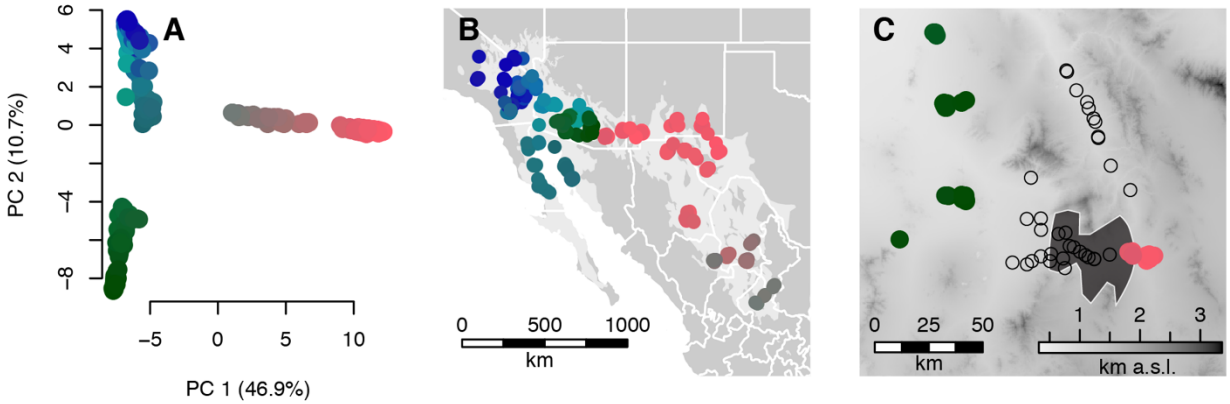

Figure 6. Range-wide genetic variation and phylogeographic breaks in *Bootettix*. **A.** PCA of 351 *Bootettix* colored by scores on PC 1 – PC 3. The first three PCs explain 62.1% of variance. **B.** Sample localities, colored as in panel A. **C.** Inset of panel B centered on phylogeographic break between eastern and western *Bootettix*. Empty circles are 33 localities we visited across three years and failed to find *Bootettix*.

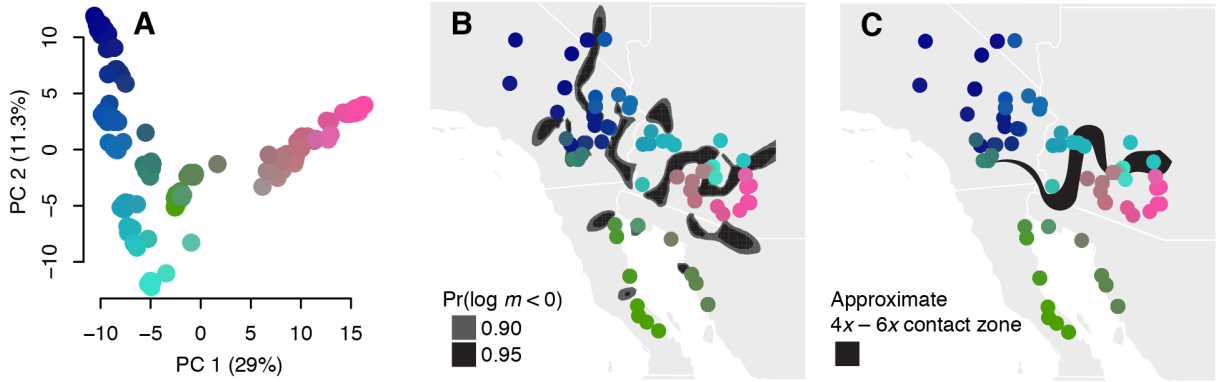

Figure 7. Genetic variation and barriers to migration in western *Bootettix*. **A.** PCA of 202 individuals colored by scores on PC1 – PC3. The first three PCs explain 48.6% of variance. **B.** Western *Bootettix* localities (colored as in panel A) mapped over barriers to migration inferred by EEMS. Barrier color represents posterior probability that migration is locally reduced below the range-wide mean (scaled to mean $m = 0$). **C.** Western *Bootettix* localities mapped over approximate path of 4*x* – 6x contact zone.

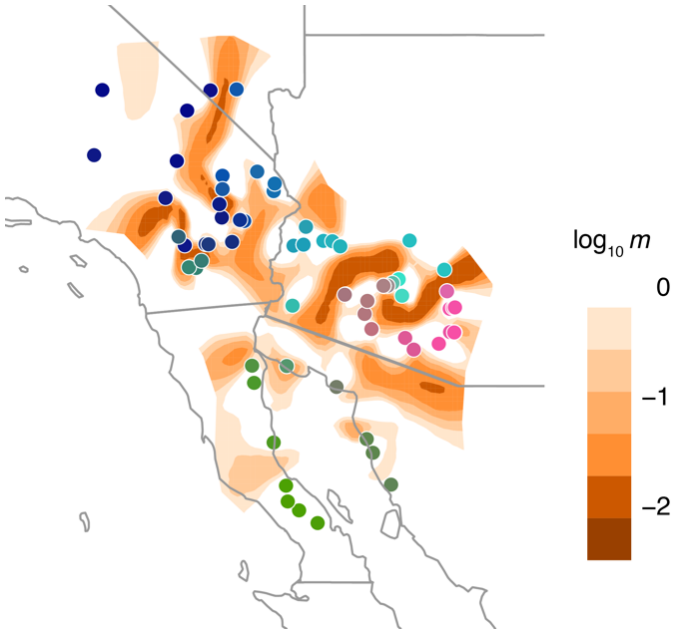

Figure 8. Strength of migration barriers in western *Bootettix* inferred by EEMS (scaled to mean $m = 0$).

Evidence of restricted migration at host plant contact zones: *Insara*

PCA of *Insara* revealed much less spatial structure than in *Bootettix* (PC1 – PC3 explained 10.1% of variance, Fig 9A). PC1 captured local population subdivision within and near the Madrean Archipelago, while PC2 separated populations from the Chihuahuan desert from all others and PC3 isolated samples from Baja California. Only the Madrean Archipelago and coincident $2x - 4x$ contact zone were significant barriers to gene flow (Fig. 9B), and these were weaker than in *Bootettix*: maximum F_{ST} across the Madrean Archipelago was 0.14.

To determine whether host plant contributes to population structure near the Madrean Archipelago, we further explored genetic turnover in southeastern Arizona. We found a cline in the major axis of genetic variation (PC1) that closely coincided with the $2x - 4x$ contact zone (Fig. 9C-D). This coincidence was remarkable given that PC1 scores were nearly invariant for > 800 km to the west.

In contrast to the fine-scale population structure near the Madrean Archipelago, we found virtually no spatial structure across most of the species' range from the Madrean Archipelago to the northwestern Mojave Desert (Fig. 9B). We therefore lack evidence for gene flow barriers near the $4x - 6x$ contact zone.

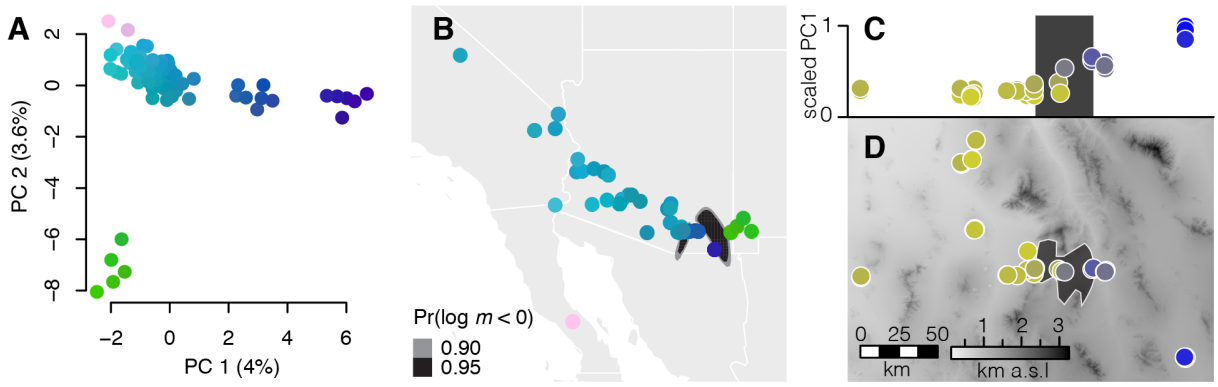

Figure 9. Genetic variation and barriers to migration in *Insara*. **A.** PCA of 115 individuals colored by scores on PC1 – PC3. The first three PCs explain 10.1% of variance. **B.** *Insara* localities (colored as in panel A) mapped over barriers to migration inferred by EEMS. Barrier color represents posterior probability of barrier location. **C.** PC1 scores (rescaled from 0-1) versus longitude. Geographic location corresponds to map in panel D. Color indicates PC1 score. Gray box shows approximate location of 2*x* – 4*x* contact zone. **D.** Map of *Insara* samples collected near 2*x* – 4*x* contact zone. Points colored as in panel C.

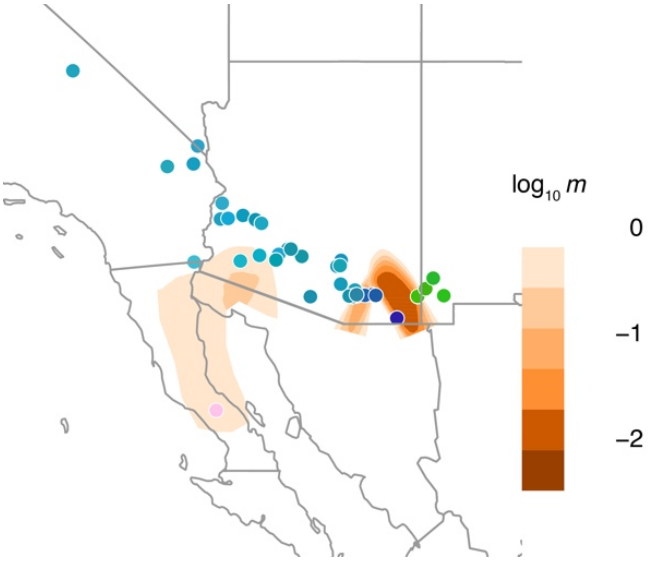

Figure 10. Strength of migration barriers in *Insara* inferred by EEMS (scaled to mean $m = 0$). Note that although EEMS modeled reduced migration near the Gulf of California, this barrier was not significant (posterior probability of barrier location < 0.9, Fig. 9B).

Geographic barriers and population structure

As noted above, the Madrean Archipelago was a prominent barrier to gene flow for both *Bootettix* and *Insara*. We identified no other genetic disjuncts between *Insara* populations, which had only weak spatial structure across the Sonoran and Mojave Deserts.

EEMS identified two additional migration barriers in eastern *Bootettix* that coincided with eastern Madrean mountain ranges (Fig. 11C). Other migration barriers in the Northern Chihuahuan Desert coincided with the Deming Plains and a mountain corridor extending from the Sacramento through the Chisos Mountains in western Texas. To the south, a migration barrier separating extreme southern populations coincided with the Southern Coahuila Filter Barrier (Fig. 11B-C). Migration was not impinged across the Pecos River.

The location of EEMS barriers suggested that gene flow was restricted between adjoining populations near major geographic features in the northern Chihuahuan Desert. If population divergence occurred *in situ*, we would expect a pattern of isolation by distance that is exaggerated across migration barriers. In fact, *Bootettix* population structure in the Northern Chihuahuan Desert was poorly correlated with geographic proximity. For example, a population within the Madrean Archipelago in Arizona was more closely related to populations in western Texas (580 km away; green in Fig. 11C) than adjoining populations in New Mexico (52 km away, deep blue in Fig. 11C). This instead suggests that pre-existing population subdivisions from the southern Chihuahuan Desert were geographically reassorted in the north during post-glacial range expansion. Small and geographically isolated founder populations would have been subject to increased genetic drift that furthered their differentiation. Consistent with this hypothesis, heterozygosity was lowest in the northern Chihuahuan Desert, where Pleistocene glaciation would have most drastically reduced desert habitat (Fig. 12A). PCA axes that show east-west geographic structure in the southern Chihuahuan Desert are geographically intermixed in the north (Fig. 12B), providing further support for this hypothesis.

Western *Bootettix* populations were highly dissected by migration barriers, many of which coincided with prominent geographic features. EEMS identified a migration barrier roughly coincident with the Colorado River (Fig 8B), although a lack of samples next to the river poorly constrained the location of this barrier. Another barrier extended longitudinally from Nevada through California along the New York, Providence, and Granite Mountains in the Mojave Desert. Creosote bush is restricted to narrow ribbons of low-elevation habitat in this mountain corridor and was likely absent during Pleistocene glaciation, which may have bisected populations. Several other mountain ranges coincide with localized migration barriers, including the Transverse Ranges in Southern California, Mesa Huatamote in Baja California, and the Sierra Madre Occidental and the Ajo Range near the Arizona-Sonora Border.

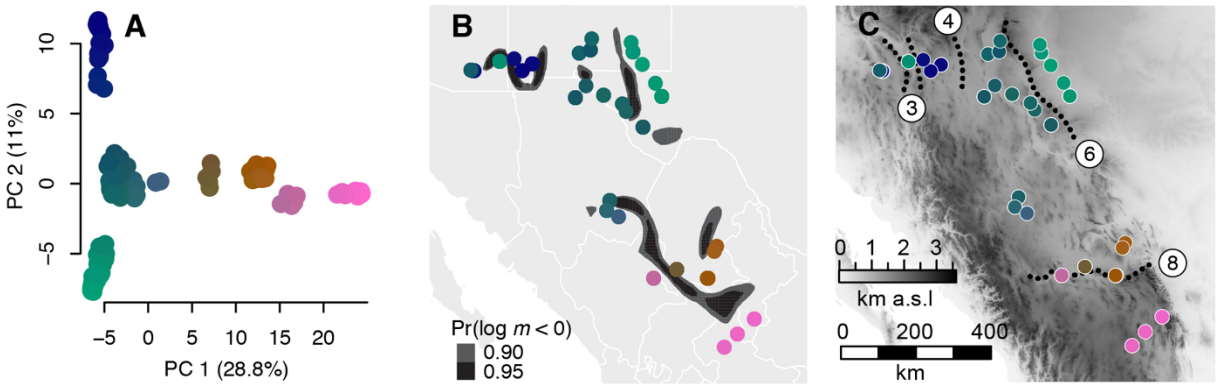

Figure 11. Genetic variation and barriers to migration in eastern *Bootettix*. **A.** PCA of 152 individuals colored by scores on PC1-PC3. The first three PCs explain 49.5% of variance. **B.** Eastern *Bootettix* localities (colored as in panel A) mapped over barriers to migration inferred by EEMS. Barrier color represents posterior probability that migration is locally reduced below the range-wide mean (scaled to mean $m = 0$). **C.** Eastern *Bootettix* localities mapped over topographical map of Chihuahuan Desert region. Biogeographic features that coincide with migration barriers identified by EEMS are indicated with dashed lines. Numbers correspond to barriers identified in Fig 1. 3: Madrean Archipelago. 4: Deming Plains. 6: Sacramento / Guadalupe / Cathedral / Davis / Chisos Mountains. 8: Southern Coahuila Filter Barrier.

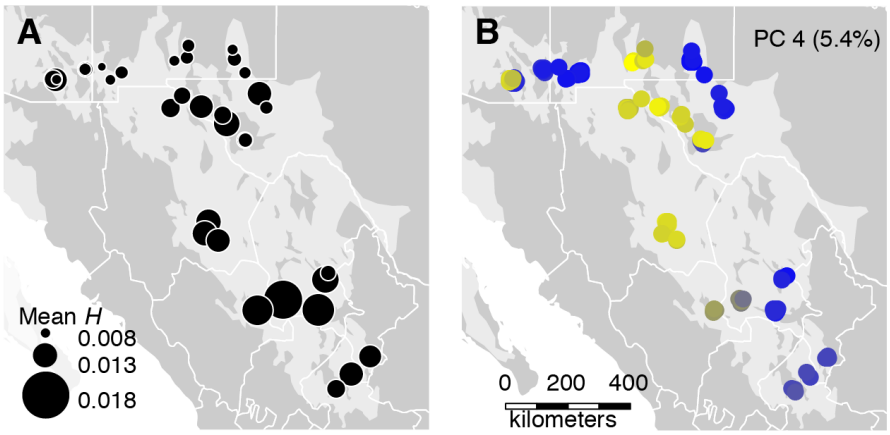

Figure 12. Genetic structure in the Northern Chihuahuan Desert is the result by recolonization of populations with existing subdivision. **A.** Mean heterozygosity was lowest in northern limits of the Chihuahuan Desert. **B.** PC axis with east-west geographic structure in the southern and central Chihuahuan Desert were geographically intermixed in the North.

DISCUSSION

Overview

The relative effects of geographic and biotic barriers to migration, and the consistency of these barriers across ecologically comparable species, is an open question. We explored population structure in two co-distributed herbivores with shared host plants in order to 1) quantify the contributions of geographic and environmental barriers to populations structure, and 2) compare these effects across species.

We found that although geography and host plant contribute to population structure in both herbivores we investigated, the magnitude of these effects was much greater in *Bootettix* than in *Insara*. Cytotype contact zones were among the strongest migration barriers for *Bootettix*, matching or exceeding well-known physical barriers in North American Deserts. In contrast, *Insara* populations were weakly differentiated across the species' range. Although we detected a small effect of host plant on range-wide patterns of genetic diversity in *Insara* and localized genetic turnover at the 2*x* – 4*x* contact zone, these effects were modest and exceeded by the effect of geography. Consequently, the pattern and strength of population subdivision differed markedly between herbivore species.

Host-associated divergence in *Bootettix*

We found that *Bootettix* comprises three major lineages associated with different creosote bush cytotypes. The especially high divergence across the 2*x* – 4*x* contact zone (F_{ST} = 0.81) and reciprocal monophyly of lineages supports the historical recognition of two *Bootettix* species: *B. argentatus* (Bruner, 1889) on diploid creosote bush in the east and *B. punctatus* (Scudder, 1890) on polyploid creosote bush in the west. Because these lineages appear to be allopatric, our data cannot directly evaluate the role of host plant in reducing migration between them.

Greatly reduced migration near the 4*x* – 6*x* contact zone evinces strong IBE for tetraploid- and hexaploid-associated *Bootettix*. Divergent natural selection or biased dispersal may explain this pattern if diverged *Bootettix* have adapted to alternative host plants (Sexton et al., 2014), as has been reported for many herbivorous insects (Forbes et al., 2017). An alternative possibility is that neutral divergence accumulated between *Bootettix* lineages in allopatry, and gene flow has only recently resumed at a zone of secondary contact near the 4*x* – 6*x* contact zone. If *Bootettix* populations were once physically isolated by lack of available habitat, we would expect to find a comparable signature of divergence in *Insara covilleae* as well. This is what we observed in the Madrean Archipelago, where a common barrier left a common signature of divergence. In fact, we found no such structure in *Insara* near the 4*x* – 6*x* contact zone, and we are unaware of any other Sonoran Desert species with similar population structure. Moreover, the long and undulating contour of the 4*x* – 6*x* contact zone, which follows no geographic features or known abiotic gradients, makes it unlikely for the *Bootettix* contact zone to coincide by chance.

Patterns of genetic divergence in *Bootettix* were therefore consistent with host-associated differentiation, which may drive eventual speciation (Nosil, 2012). (We avoid the term "host race" due to its emphasis on divergence in sympatry, Drès & Mallet, 2002). Host-associated divergence is a common phenomenon (Feder et al., 1988; Abrahamson & Weis, 1997; Via, 1999; Nason et al., 2002; Stireman et al., 2005; Peccoud et al., 2009; Forbes et al., 2017), although by no means ubiquitous (Stireman et al., 2005; Forbes et al., 2017). Examples reported to date involve

divergence between populations using distantly related (e.g., Funk et al. 2011) or congeneric hosts (e.g., Stireman et al. 2005), but not sister species. Against this backdrop, the *Bootettix* – creosote bush system is particularly interesting because of the possibility that *Bootettix* has co-diverged with its hots plant through successive autopolyploidy events. Co-divergence of insects and host plants has been rarely reported in herbivorous insects (Farrell & Mitter, 1990; Farrell, 2001; Smith et al., 2008, 2011). In the future, demographic modeling of creosote bush and *Bootettix* may resolve whether their divergence times are coincident, as expected under a scenario of co-divergence.

Causes of incongruence between herbivore species

The modest genetic differentiation in *Insara* was striking in light of the substantial divergence between *Bootettix* populations. Several factors may explain this incongruence between species. First, dispersal distance may differ between *Bootettix* and *Insara*, which has previously been shown to predict population structure (Bohonak, 1994), including in other desert insects (Phillipsen et al., 2015). Long-range dispersal has also been invoked to explain limited population structure across for other katydid species (Ney & Schul, 2017).

While we have no data on dispersal of either herbivore we studied, field observations and natural history reports indicate that *Insara* is more dispersive than *Bootettix*. *Insara* are commonly attracted to artificial lights (Tinkham 1938, pers. obs.), suggesting that that they are active night-time fliers. Although *Bootettix* have also been anecdotally reported from lights, in our experience this is rare. Long-range dispersal in *Insara* would reduce IBD, permit migration over or around barriers that affect less dispersive species, and impede host plant adaptation by homogenizing populations near contact zones. Our finding that *Insara* was modestly differentiated across the 2*x* – 4*x* contact zone (where migration is limited to a 14 km-wide valley) but not along the 4*x* – 6*x* contact zone (where migration occurs across a long, spatially complex interface) is consistent with this hypothesis.

Second, phenotypic differences between creosote bush cytotypes that are salient to *Bootettix* may be irrelevant to *Insara*. Such differences could include levels of a major defensive compound found in creosote bush resin (Zuravnsky, 2014), volatile organic products (Bohnstedt & Mabry, 1979) that may serve as host-finding cues, or host plant phenology (Laport et al., 2016). Depending on the dynamics of plant-insect interactions for each herbivore, such differences may or may not impose divergent selection or affect habitat choice. Experimental investigation of interactions between creosote bush and its herbivores will be required to evaluate these hypotheses.

The effect of polyploidy on species interactions

Our results are consistent with a growing body of evidence that cytotype variation within nominal plant species can affect interactions with herbivores (Thompson et al., 1997, 2004; Segraves & Thompson, 1999; Nuismer & Thompson, 2001; Janz & Thompson, 2002; Halverson et al., 2008; Kao, 2008; Arvanitis et al., 2010; Münzbergová et al., 2015; O'Connor et al., 2019). In the case of *Bootettix*, host plant polyploidy may even trigger concomitant divergence in herbivore populations, thereby linking plant and insect diversification. We conclude that plant polyploidy can have broad effects on the ecology, evolution, and biogeography of plant-associated communities (Segraves & Anneberg, 2016; Laport & Ng, 2017; Segraves, 2017; O'Connor et al., 2019).

CHAPTER 3: Unresolved sexual antagonism maintains an ancient polymorphism in the desert clicker grasshopper

Anticipated co-authorship: Jiarui Wang, Marissa Sandoval, and Noah K. Whiteman[1]
1. Department of Integrative Biology, University of California Berkeley, Berkeley, CA 94706

ABSTRACT

The processes that maintain genetic variation remain an enduring fascination in evolutionary biology. Classic forms of balancing selection (overdominance, negative frequency dependence, spatially or temporally variable selection) can sometimes maintain adaptive polymorphisms for millions of years. Opposing selection between sexes – sexual antagonism (SA) – is a less-studied form of balancing selection that makes unique predictions about the genomic location and fate of balanced polymorphisms. Theory predicts that autosomal polymorphisms balanced by SA are unlikely unless selection is strong ($s \geq 0.2$) or there are opposite dominance relationships in each sex. Sex-linked loci may also be maintained over a broader range of parameters. Over time, selection should favor the resolution of SA through sexual dimorphism, which may intrinsically limit the duration of balancing selection by SA. However, these predictions have not yet been tested with a system that links the context of phenotypic selection with the evolutionary history of associated loci. To address this gap, we studied SA on a crypsis polymorphism in the desert clicker grasshopper (*Ligurotettix coquilletti*). Crypsis is in this species is defined by color, which varies quantitatively, and a discrete pattern morph (either banded or uniform). Males are exposed to predation only on host plant stems, while females are most vulnerable while laying eggs on the ground. SA may thus arise if the predation on the ground vs. stems selects for different phenotypes. To understand the context of selection, we quantitatively compared the color and pattern (defined by light-dark contrast over a given spatial scale) in grasshoppers and their environments. We found that while uniform morphs resemble stem color, banded morph frequency was positively correlated to the strength of local ground patterning. Phenotypes did not differ between males and females, indicating no progress towards resolving SA. With RADcap sequencing and association mapping we identified a single Mendelian, autosomal locus associated with pattern morph. Surprisingly, we found that divergent alleles at this locus were shared between *L. coquilletti* and its sister species (trans-species polymorphism), which has a similar pattern polymorphism. Our results indicated that SA can maintain polymorphism for perhaps millions of years. We speculate that the longevity SA may be due to an interaction with other forms of balancing selection that act solely on females.

INTRODUCTION

The processes that maintain genetic variation remain an enduring fascination in evolutionary biology. While a substantial fraction of polymorphism is selectively neutral or nearly so (Ohta, 1992), there is also abundant evidence of polymorphism at loci that affect fitness (Charlesworth, 2006). How adaptive variation is maintained in the face of drift and directional selection was an early quandary in population genetics. As such, natural selection that preserves genetic variation (balancing selection) has been intensively studied from both theoretical and empirical perspectives.

Overdominance was the first form of balancing selection to be articulated (Fisher, 1922) and was the focus of most early research on the topic (Hedrick, 2012). In a textbook example, the allele

that causes sickle-cell anemia when homozygous confers protection against malaria when heterozygous, rendering heterozygotes most fit where malaria is common (Allison, 1954). While this example is instructive, overdominance is unlikely to affect many sites in the genome due to the genetic load in imposes (Haldane 1957; Lewontin and Hubby 1966; but see Agrawal and Whitlock 2012). Available evidence suggests overdominance is a relatively uncommon form of balancing selection (Hedrick, 2012).

More common forms of balancing selection arise when the strength and/or direction of selection is context-dependent. Selective environments that vary in space (Levene, 1953) or time (Dempster, 1956) can maintain alleles that confer an advantage in some contexts, but disadvantage in others. Both forms of selection have been demonstrated in the lab (Huang et al., 2014) and field (Bergland et al., 2014). Negative frequency-dependent selection, which affords an advantage to rare alleles, underlies polymorphism at self-incompatibility loci in plants (Delph & Kelly, 2014), sexual selection in a variety of vertebrates (e.g., Sinervo and Lively 1996), and is a common feature of loci mediating antagonism between species (e.g., Hori, 1993; Bergelson et al., 2001; Bakker et al., 2006; Bond, 2007; Karasov et al., 2014).

Balancing selection is likely to be a transient phenomenon in most cases (Charlesworth, 2006; Fijarczyk & Babik, 2015), but under some conditions it may persist for millions of years (Gao et al., 2015). A hallmark of ancient balancing selection is trans-species polymorphism, where the balanced alleles found in multiple species are more ancient than the species that carry them (Clark, 1997; Wiuf et al., 2004). Only a handful of cases of trans-species polymorphism are known (Figueroa et al., 1988; Uyenoyama, 1997; Städler & Delph, 2002; Surridge & Mundy, 2002; Ségurel et al., 2012; Leffler et al., 2013; Van Diepen et al., 2013; Rasmussen et al., 2014). Examples to date are mostly cases of negative frequency-dependent selection on loci implicated in host-parasite interactions or reproduction (self-incompatibility, cytoplasmic male sterility). However, it is unclear to what extent other forms of balancing selection, and selection on different organismal traits, can result in long-term polymorphisms (Gao et al., 2015).

An additional form of balancing selection – sexual antagonism – has received less empirical attention. Intra-locus sexual antagonism (hereafter, SA) arises when male and female traits are positively correlated but divergently selected between sexes (Bonduriansky & Chenoweth, 2009; van Doorn, 2009). The first empirical evidence of SA came from negative correlations between male and female fitness in *Drosophila melanogaster* (Rice & Chippindale, 2001), a finding that has been since been replicated in a number systems (e.g., Foerster et al., 2007). While reproductive traits were an early focus of research on SA, traits that affect any fitness component (e.g., survival) may be under sexually antagonistic selection. There are now many examples from the lab and field (Cox & Calsbeek, 2009), and theory predicts that SA is an inevitable outcome of evolution in dioecious species (Connallon & Clark, 2014a, 2014b).

The dynamics of SA make several unique predictions about the genetic architecture of traits under balancing selection and their evolutionary trajectory. Although classic theory established SA as a form of balancing selection (Owen, 1953; Haldane, 1962; Kidwell et al., 1977), stable polymorphism at an autosomal locus requires nearly equal and opposite selection in each sex when selection is weak-to-moderate ($s \leq 0.2$) (Kidwell et al., 1977; Patten & Haig, 2009). Alternative genetic architectures can be more permissive. Sex-linkage can expand the conditions permitting

polymorphism (Kidwell et al., 1977; Rice, 1984; Jordan & Charlesworth, 2012), but not always (Curtsinger, 1980; Vicoso & Charlesworth, 2006; Fry, 2010). Dominance reversals – in which allelic dominance differs between the sexes in a way that enhances fitness (Gillespie, 1978) – makes balancing selection by SA still more attainable (Fry, 2010; Jordan & Charlesworth, 2012; Connallon & Chenoweth, 2019). Such dominance reversals have recently been recently found in salmonid fish, which lack heteromorphic sex chromosomes (Barson et al., 2015; Pearse et al., 2019). In light of empirical examples of sex-linked and autosomal polymorphisms under SA selection (Fry, 2010), the genomic architectures that permit SA polymorphisms in nature awaits further investigation.

Because SA displaces both sexes from their respective fitness peaks (Long et al., 2006; Connallon et al., 2010; Matthews et al., 2019), selection is predicted to resolve sexual antagonism over time (Lande, 1980). A deterministic march toward resolution may thus intrinsically limit the duration of balancing selection by SA, although this has not been examined empirically. Mechanisms of conflict resolution include sex-linked modifier loci that break male-female genetic correlations (which underliies the fitness cost of SA) (Vicoso & Charlesworth, 2006; Bonduriansky & Chenoweth, 2009; van Doorn, 2009). With time, the genetic architecture of antagonistically selected traits is expected to evolve in ways that permit sexual dimorphism (Lande, 1980; Bonduriansky & Chenoweth, 2009).

While there is strong evidence that SA is pervasive and a potentially common form of balancing selection, there is no empirical evidence regarding the evolutionary trajectory of alleles under SA. This represents a notable gap in our understanding of balancing selection.

The desert clicker grasshopper (*Ligurotettix coquilletti*) presents an opportunity to close this gap. Desert clicker crypsis is composed of two traits (pattern and color) that vary within and among populations in the Sonoran, Mojave, and Peninsular Desert of North America. Two discrete pattern morphs typically coexist within populations. Uniform morphs have relatively homogeneous color (Fig. 1A-C), while banded morphs have contrasting light and dark bands along the body axis (Fig. 1 D-F). The absence of intermediate forms suggests the pattern polymorphism is determined by a single Mendelian locus. Independent of pattern, color varies quantitatively within and among populations. Uniform morphs range from ivory (Fig. 1A) to deep brown-gray (Fig. 1C), and the contrast of banding can also be strong (Fig. 1D) or almost masked by dark pigment (Fig. 1F) in banded morphs. A similar polymorphism is expressed in a recently-discovered sister species of *L. coquilletti* (*L.* sp. "bajaensis", Fig. 2; T.K. O'Connor unpubl. data).

SA may arise in *L. coquilletti* due to life history differences between the sexes. Males maintain territories on stems of their primary host plant (creosote bush, *Larrea tridentata*) and seldom leave (Wang & Greenfield, 1994). While calling from plant stems to ward off rivals, males also expose their position to predators. Females are typically silent and spend most of the day on host plant stems. However, they must descend to the ground in the early morning to oviposit, where they are immobilized for $\geq$ 45 minutes (Wang & Greenfield, 1994) at a time that coincides with peak vertebrate predator activity. Each sex is therefore more vulnerable to predation in a different environment. Because the visual properties of these environments differ, alternative cryptic phenotypes may be favored in each sex.

Figure 1. Crypsis variation in the desert clicker (*Ligurotettix coquilletti*). Each row is a pair of uniform (left) and banded (right) males collected from the same site.

It was not obvious *a priori* if or how environmental variation might favor particular crypsis phenotypes. Color matching is a common form of crypsis, so grasshoppers might match the color of the environment where they are depredated. Independent of color, variation in environmental patterning – localized alternation between light and dark – may also influence detectability. Homogeneous prey in a heterogeneous habitat may be more readily detected than prey that match the strength and scale of patterning in their environment. Creosote bush stems and the ground at our study sites varied substantially in both color and pattern. Stems ranged from blanched gray to charcoal, with varying shades of lighter red- and yellow-brown. While some stems have a homogeneous color, others had dark resinous blotches at regular intervals created strong contrasting patterns. Likewise, the granitic, volcanic, and sandy soils of desert environments differed in their color, brightness, granularity, and homogeneity. There was thus ample variation in both color and pattern of each environment that may affect predation risk.

Here, we explored the ecological context of selection on crypsis variation and the evolutionary history of associated loci. We first quantitatively characterized distinct aspects of grasshopper crypsis (color and pattern morph) to determine whether males and females express similar phenotypes. We then used phenotype-environment correlations to identify the ecological context of selection and how it might differ between sexes. With RADcap sequencing (Hoffberg et al., 2016) and association mapping, we next identified a single locus strongly associated with pattern morph. Our data allowed us to determine the genomic location (autosomal vs. X) and dominance of alleles at this locus for each sex. We then investigated the evolutionary history of the pattern-associated locus to characterize the persistence of polymorphisms balanced by SA.

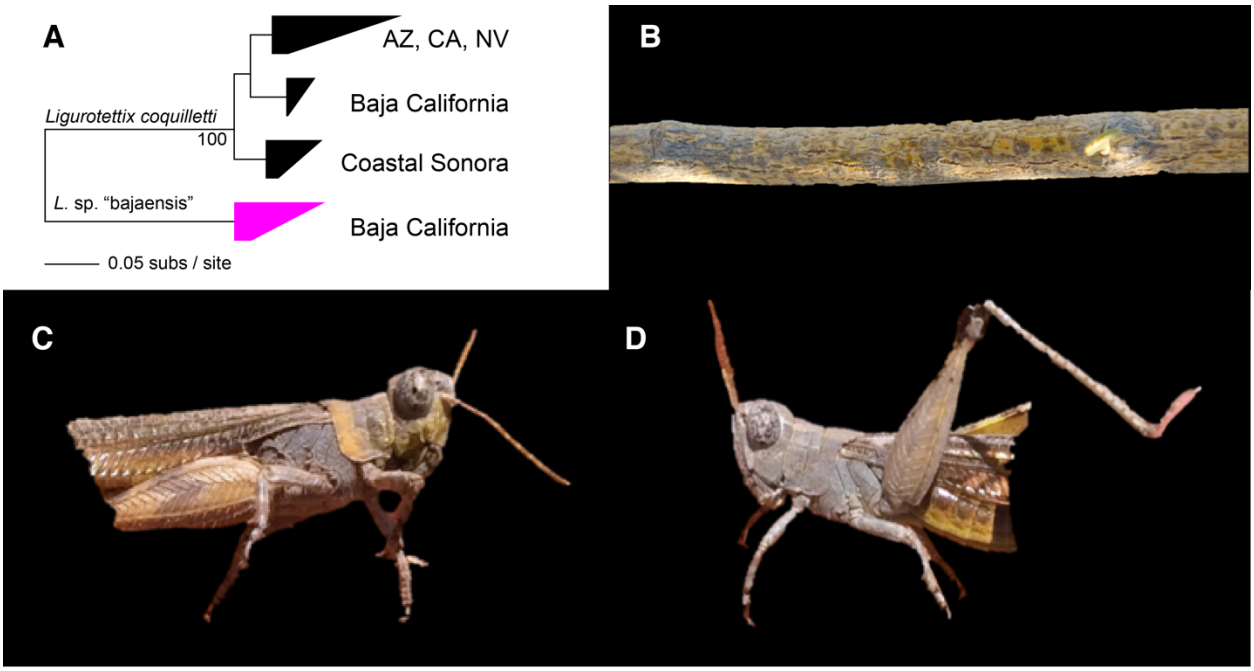

Figure 2. The Baja clicker (*Ligurotettix* sp. "bajaensis") expresses a banding polymorphism similar to *L. coquilletti*. **A.** Simplified Maximum likelihood phylogeny of 183 individuals collected from across the range of *L. coquilletti* (based on 583 unlinked SNPs). Eight individuals collected in Baja California form a well-supported clade that is distinct from *L. coquilletti* even in sympatry. **B.** Most (7/8) of the *L.* sp. "bajaensis" individuals were collected from *Viscainoa geniculata*. Stem pictured here. **C**. Banded morph of *L.* sp. "bajaensis", with yellow / gray pattern. **D.** Uniform morph of *L.* sp. "bajaensis". Both banded and uniform morphs of *L.* sp. "bajaensis" resemble the gray and yellow stems of *Viscainoa geniculata*.

METHODS

COLLECTIONS

Primary field surveys

We surveyed grasshoppers and their environment at 16 sites in Arizona and California in August-September, 2018 (Fig. 3). At each site we collected *L. coquilletti* (median n = 17, range 4 – 23) and noted the banding phenotypes of grasshoppers that we observed but did not capture (total observations per site: median = 24, range = 4 – 47). In total we observed 390 grasshoppers and captured 260. We also collected six 15-cm stem clippings from each of 10 creosote bush individuals separated by ≥ 5 m (three 6 mm diameter stems, three 12 mm diameter stems), which we later photographed in the lab.

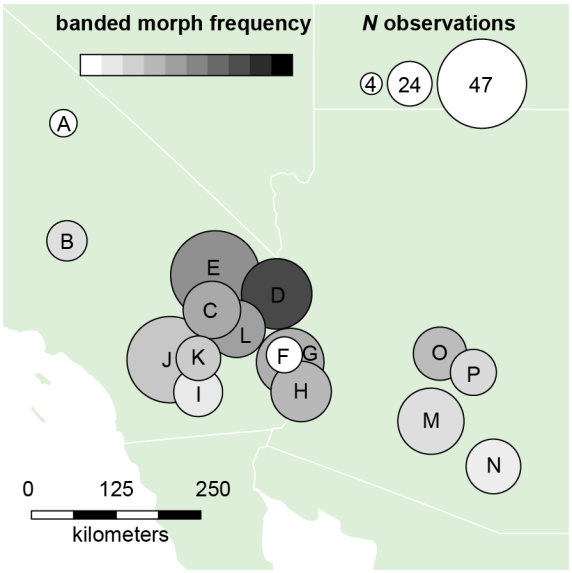

Figure 3. Map of survey sites indicating the number of grasshopper observations and local frequency of the banded morph.

Photography

All photographs were captured in RAW format with a stock Nikon D7500 camera mounted on a horizontal bracket at a height of 60 cm. At the start of each imaging session we photographed an XRite ColorChecker and scale bar under the same lighting conditions used for photographing grasshoppers and their environment. We later used these images to color-correct and rescale photographs, permitting standardized comparisons of pattern and color among all images.

We used a lightbox fashioned from a 40 cm^3 insect cage and off-camera flashes to evenly illuminate images. Two opposing vertical faces of the lightbox were covered with polyester batting, which transmitted and diffused light from two Yongnou 560 Speedlites. Lights were fitted with gel diffusers and oriented at 15˚ above horizontal. The other vertical faces were covered in reflective white foamboard. The top face of the lightbox was also covered with foamboard, with a 10 cm x 15 cm hole to permit overhead photography. The bottom face was open so the lightbox could be easily placed over subjects.

Female *L. coquilletti* lay eggs in the rocky substrate between creosote bushes, typically 1-3 m from the nearest plant (Wang & Greenfield, 1994). We therefore photographed a ~20 cm x 30 cm patch of ground 1 m from the base of ≥10 host plants per site to represent likely predation environment. Photos were captured with a 40mm Nikkor Micro f/2.8 lens at f/22 (to maximize depth of field). We captured a total of 193 images of the ground.

A total of 917 stems were photographed in the laboratory against a low-reflection black velvet background. All stems were stored cool conditions for up to two days until photographed. Photos were captured with a 40mm Nikkor Micro f/2.8 lens at f/10.

In order to preserve grasshopper phenotypes, we maintained live grasshoppers in captivity until imaging (1-3 days). We immobilized grasshoppers by chilling immediately before photographing them in lateral view against a black velvet background with a Tamron 90mm f/2.8 lens at f/10.

Secondary collections

For molecular analyses, we included an additional 290 individuals collected from across the full distribution of *L. coquilletti* between July 2015 and September 2018 (Figure 4). Of these, 202 were scored as banded or uniform.

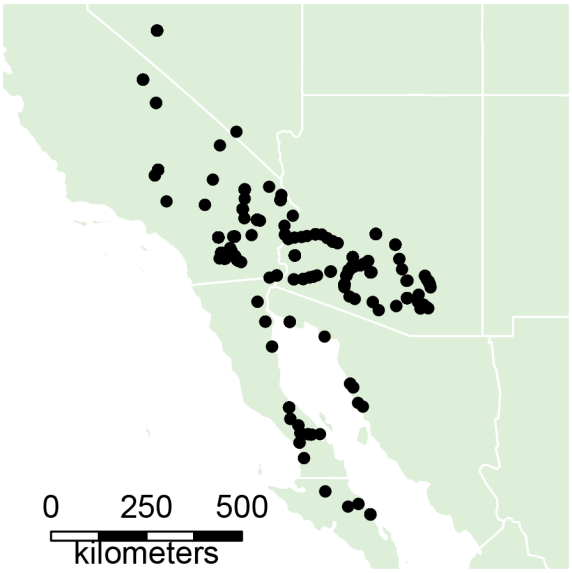

Figure 4. Map of collections used for DNA sequencing. This covers nearly the full extent of *L. coquilletti* (populations are also known from southern Baja California)

IMAGE ANALYSIS

Pattern analysis

Image processing

We used the Mica Toolbox v1.22 (Troscianko & Stevens, 2015) in ImageJ (Abràmoff et al., 2004) to color-correct and rescale RAW images, select regions of interest, and calculate image luminance under a model of human vision. While this luminance channel does not include UV and does not directly model predator perception, it provided a first-order approximation of light-dark variation in grasshoppers and the environments they inhabit.

Selecting regions of interest

We found it was not possible to position grasshoppers in poses that were consistent across all individuals. In particular, bright yellow-orange abdominal coloration that is usually obscured by the legs of resting grasshoppers was visible to varying degrees. We therefore selected two body regions that were consistently visible in all images and effectively captured color and pattern variation among grasshoppers: pronotum and femur. All further analyses were conducted using only those regions to represent grasshopper phenotypes. Results were qualitatively similar for both body parts, so we report only analyses using pronotum pattern and color.

We manually isolated stem images from their background and removed portions of the image that reflect damage due to processing (e.g., bare wood at clipped ends of stem). Ground images were cropped to a 15 cm^2 square in the center of each image.

Pattern quantification

We used fast Fourier bandpass filtering (Stoddard & Stevens, 2010) to analyze pattern energy – that is, the magnitude of light-dark alternation over a defined spatial interval – at scales ranging from 2 to 256 pixels, with a multiplicative step size of $1+1^{1/3}$. For grasshoppers this corresponded to spatial a range of 0.038 – 4.92 cm, while for stems and ground this corresponded to 0.12 cm to 14.77 cm. This analysis also quantified overall image contrast, defined as the standard deviation of luminance.

Color analysis

Image processing

Our color quantification workflow required images in PNG format (see below). We therefore converted RAW files to PNGs via a JPEG intermediate, as a direct RAW to PNG converter was unavailable. To do so, we first exported RAW images to JPEG using Photoshop CC 2019. We then used ImageJ to isolate the same regions of interest used for pattern analysis, superimpose these regions over a pure green background (aiding in subject-background discrimination in later processing steps), and save them in PNG format. We used Photoshop rather than ImageJ for the initial JPEG conversion because we noticed that ImageJ output JPEGs with substantial color noise and compression artifacts even at the highest quality setting. We found no such issues using Photoshop, and PNGs generated by ImageJ were not degraded.

We calibrated the color of resulting PNG files with images of an XRite ColorChecker and a modified colorchecker command from the 'patternize' package (Van Belleghem et al., 2018) in R (R Core Team, 2019). Our custom version of this command allows input and output of PNG files (only JPEG accepted in published implementation). Code is available from the authors.

Color quantification

We implemented our color quantification workflow with the 'colordistance' package (Weller & Westneat, 2019) in R. First, we summarized the distribution of colors in each color-calibrated image by binning pixels based upon red, green, and blue (RGB) values. Each color axis was divided into 5 bins, resulting in $5^3 = 125$ RGB color bins. Pure green background pixels were excluded from binning. We next quantified color differences between all images with the earth mover's distance (EMD, Rubner et al. 2000). EMD quantifies the minimum amount of "work" required to transform one color distribution to another, considering both the number of pixels to be moved among bins as well as the distance between bins in RGB space. We performed a principal coordinates analysis (PCoA) upon the EMD matrix to visualize color variation among all images.

HYPOTHESIS TESTING

Evaluating sexual dimorphism

We used Fisher's exact tests to determine whether morph frequencies differed between sexes. We considered all sites where we observed ≥ 4 females (10 sites), as well as an additional site from Nevada surveyed in September 2018 (38.51312˚ N, -118.0737˚ W).

We next tested whether the expression of each morph phenotype differed between males and females. Banded and uniform morphs are differentiated both on the degree of contrast and the organization of contrasting colors into well-defined bands. If selection has favored partial resolution of SA, we predicted that one or both of these quantities would differ between sexes for one or both morphs. We therefore used a Gaussian linear model to test the effect of sex on 1) overall contrast, and 2) banding strength, defined as summed pattern energy from 0.5-2 cm (the spatial scale at which banding is expressed in both sexes). Because phenotypes may also scale allometrically and differ among sites, we included a proxy body size (image area) and site as covariates. We evaluated the significance of each predictor variable with an ANOVA (type II sum of squares) using the Anova function in the 'car' package (Fox & Weisberg, 2019) in R.

Finally, we tested whether the color of each morph differed between sexes and among sites using a permutational MANOVA (PERMANOVA, Anderson 2001) implemented with 'vegan' package (Oksanen et al., 2017) in R. We extracted the marginal effects of each term with the adonis2 function.

Environment-phenotype correlations: morph frequencies

We first used binomial generalized linear mixed-effect models (GLMMs) to test whether the proportion of banded individuals depended upon the patterning of stems, patterning of the ground, or their interaction (with site as a random effect). We separately tested for an association between morph frequencies and environmental patterning that matched the scale of grasshopper banding patterns ("band-scale patterning", 0.5 cm – 2 cm), smaller-scale patterning (0.1 cm – 0.4 cm), and larger-scale patterning (2.5 cm – 5 cm). We evaluated the significance of model terms using likelihood ratio tests implemented with the Anova function in 'car'.

Environment-phenotype correlations: grasshopper color

We reasoned that if each morph was favored in a different environment, the degree of color similarity to stems or the ground might differ between morphs. We therefore used linear mixed-effect models (LMMs) to test the hypothesis that the color of each morph was unequally similar to stems vs. the ground. To construct our models, we first summarized the mean color distance (EMD) between each grasshopper and 1) all stems, or 2) all ground images. For each color morph we separately modeled log-transformed EMD as a function of comparison type (stems or ground) with a random effect of individual to account for the paired nature of the data. We evaluated the significance of model terms with a Wald *F* test implemented with the Anova function in the 'car' package in R.

As a stronger test for color matching, we then asked if grasshopper color more closely resembled environments at the site where they were captured versus environments from different sites (local adaptation). Again, we first summarized the mean EMD between each grasshopper and local environments (within-site comparisons) or non-local environments (between-site comparisons). For each environment (i.e., stem or ground), we then used LMMs to model log-transformed EMD as a function of comparison type (within or between sites), with a random effect of individual nested within site to account for the paired nature of the data. As above, we evaluated the significance of model terms with a Wald *F* test.

MOLECULAR METHODS

Sequence capture design

Our basic strategy for designing sequence capture baits follows methods described in Chapter 2. Briefly, we previously performed pilot ddRAD sequencing on 28 individuals of *Ligurotettix coquilletti* collected from across Arizona, California, and Nevada. From the resulting ~385,000 loci that we assembled with ipyrad (Eaton, 2014), we identified loci that were suitable for DNA sequence capture with respect to their GC content, number of indels, and levels of sequence divergence. We designed probes for a random subset of 9,300 such loci following methods described in Chapter 2.

In addition, we attempted to use these pilot data to identify loci associated with pattern morph. The high missingness in our sequencing dataset (often < 50% of individuals genotyped per locus) limited our statistical power, making it difficult to discern true associations. However, the large number of loci we assembled afforded relatively dense coverage of the *L. coquilletti* genome, increasing the chances of detecting an association with pattern morph if one existed. Assuming a 10 Gb genome (typical of grasshoppers; Gregory 2019), 385,000 loci corresponds to 1 locus per ~26 kb. Fifteen of the 28 individuals we sequenced had been phenotyped as banded or uniform. We therefore used these 15 individuals to test for an association between pattern morph and genotype at each SNP using Fisher's exact tests implemented in R. We designed sequence capture probes for loci with the top 850 most significant associations.

As described in Chapter 2, we combined sequence probes for *L. coquilletti* with those from three other species, each representing ~10,000 probes in a pool of 40,000 myBaits Custom probes synthesized by Arbor Biosciences (Ann Arbor, MI).

RADcap library preparation and sequencing

Our general protocol for sequence capture followed the methods described in Chapter 2. We prepared RADcap libraries for all 540 individuals and sequenced them across portions of two HiSeq 4000 lanes at the Vincent G. Coates Sequencing Laboratory at UC Berkeley.

Identifying loci associated with pattern morph

We performed a reference-based RAD assembly using ipyrad v0.9.13 with default QC parameters. We used a unique molecular index in the i5 index read to remove PCR duplicates, then mapped reads against the loci used to design sequence probes. We called genotypes for loci with ≥ 6x coverage and retained loci with ≥ 20 individuals represented. As in preliminary analyses, we then used genotype calls to perform Fisher's exact tests of genotype-phenotype associations in R.

Our analysis identified a single locus strongly associated with morph phenotype (see results). However, due to poor sequencing performance, this locus was assembled in a minority of samples. One possible cause was that this locus was sequenced in other samples but to insufficient depth to be retained in the final assembly. To increase the number of genotypes available for analysis, we therefore used BLASTN to query demultiplexed, quality-trimmed sequencing reads for individuals from a subset of sites. Our search focused on 12 sites that had both high numbers of sequenced individuals (≥ 10) and a relatively high proportion of banded morphs to increase the probability of genotyping multiple individuals from each morph from within a site. We assembled reads with significant hits to the pattern-associated locus and added these to our existing alignment. In our

search we discovered that many individuals had reads matching one half of the pattern-associated locus, but not the other. This may be due to restriction-site polymorphism near the RAD locus or incomplete enzymatic digestion. As a result, the 54 additional haplotypes we identified and assembled included only half of the locus (89 bp).

Genealogy of pattern-associated locus

We combined haplotypes assembled by ipyrad with manually assembled partial haplotypes into a single alignment using MUSCLE v3.5 (Edgar, 2004). We then inferred a maximum likelihood phylogeny with RAxML (Stamatakis, 2014) under a GTR+G model within Geneious v10.2.3 (Biomatters, LTD). We assessed branch support with 200 bootstrap replicates.

Molecular evidence of balancing selection

To evaluate molecular evidence of balancing selection, we calculated Tajima's *D* based upon partial haplotypes (89 bp) at the pattern-associated locus. We focused on three sites with ≥ 10 haplotypes (range: 11-19), and calculated significant deviations from neutrality (expectation: $D = 0$) as in (Tajima, 1989).

RESULTS

No sexual dimorphism in crypsis

The proportion of banded morphs did not differ between sexes at 11 sites (Table 1). Likewise, neither banding strength nor total contrast differed between sexes for banded morphs (Table 2, Table 3). For uniform morphs, there was a slight but significant effect of sex on contrast ($R^2 = 0.016$, $P = 0.01$), but this was considerably weaker than the effects of body size or site (Table 3). Sex was not associated with color variation among banded morphs, and weakly associated with color variation in uniform morphs ($R^2 = 0.02$, $P < 0.001$, Table 3). By comparison, color variation was strongly associated with site in both banded and uniform morphs ($R^2 = 0.68$ and $R^2 = 0.55$, respectively, both $P < 0.0001$). Taken together, these results suggest that crypsis traits were not meaningfully differentiated between the sexes. All subsequent analyses of phenotypes pool males and females.

Table 1. Count of pattern phenotypes for males and females at 11 sites where we observed ≥ 4 individuals of each sex. *P* values report Fisher's exact tests. P_{FDR} are *P* values adjusted for 5% false discovery rate. Site names correspond to labels in Fig. 3. "-" indicates a site not included in primary analyses for this study.

	male		female				
site	banded	uniform	banded	uniform	**total**	***P***	P_{FDR}
C	8	5	2	10	25	0.041	0.455
D	17	3	10	4	34	0.41	1
E	22	21	2	2	47	1	1
G	10	18	2	2	32	0.62	1
H	8	15	1	3	27	1	1
I	2	12	0	5	19	1	1
J	8	25	4	8	45	0.705	1
K	2	9	2	3	16	0.547	1
N	2	15	0	6	23	1	1
P	1	12	2	2	17	0.121	0.663
-	5	8	0	4	17	0.261	0.955

Table 2. Results of ANOVA testing for an effect of sex, size, and site on *L. coquilletti* banding intensity (banded morphs only).

factor	d.f.	s.s.	R^2	*F*	*P*
sex	1	0.015	0.003	1.36	0.248
size	1	0.059	0.010	5.22	**0.025**
site	15	4.815	0.843	28.5	**2.0×10^{-16}**
residuals	73	0.823			

Table 3. Results of ANOVA testing for an effect of sex, size, and site on *L. coquilletti* contrast. Each morph is considered separately.

morph	factor	d.f.	s.s.	R^2	*F*	*P*
banded	sex	1	0.005	0.001	0.464	0.498
	size	1	0.043	0.009	4.08	**0.047**
	site	15	3.965	0.829	25.0	**2.0×10^{-16}**
	residuals	73	0.772			
uniform	sex	1	0.063	0.016	6.82	**0.010**
	size	1	0.350	0.087	38.1	**4.1×10^{-09}**
	site	19	1.877	0.468	10.7	**2.2×10^{-16}**
	residuals	187	1.721			

Table 4. Results of permutational MANOVA testing the marginal effects of sex and site on color variation among individuals. (Note: because we report marginal effects, R^2 values will not sum to 1).

morph	factor	d.f.	s.s.	R^2	*F*	*P*
uniform	sex	1	0.039	0.021	10.6	< 0.001
	site	19	1.021	0.553	14.6	< 0.0001
	residuals	188	0.692	0.375		
banded	sex	2	0.008	0.009	1.23	0.282
	site	15	0.61	0.68	11.9	< 0.0001
	residuals	74	0.252	0.281		

Banded morphs are more common where ground is strongly patterned

Grasshoppers were more likely to be banded at sites with strong ground patterning at a scale of 0.5 – 2 mm ("band-scale patterning", binomial GLMM, $b = 0.68$, $P = 1.4 \times 10^{-4}$, Fig. 5). This matches the scale of light-dark alternation in banded morphs, and was a better predictor of grasshopper morph than patterning at larger ($b = 0.56$, $P = 0.0072$) or smaller scales ($b = 0.56$. $P = 0.0049$). By contrast, grasshopper morph was unrelated to stem patterning at any spatial scale (all $P > 0.3$). Likelihood ratio tests did not support more complex models including stem or ground color in addition to stem patterning (all $P > 0.1$)

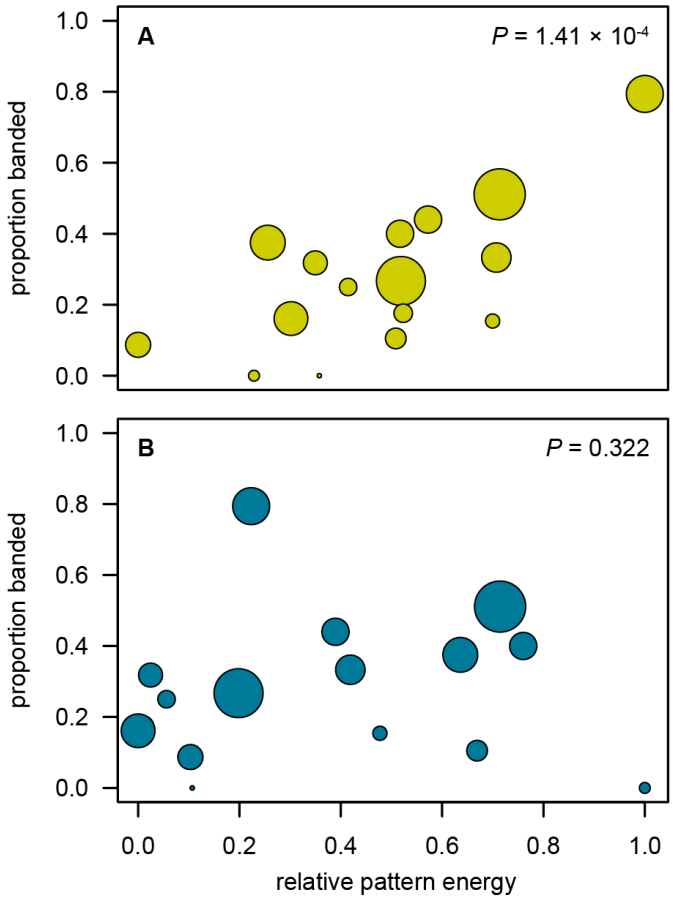

Figure 5. The frequency of banded grasshoppers is positively correlated to band-scale pattern energy in the local ground (**A**) but not local stems (**B**). Pattern energy rescaled from 0-1.

Pattern morphs are color-matched to different environments

A PCoA of pairwise color distances separated grasshoppers and their environments along two major axes (Fig. 6). The second axis (explaining 19.1% of variation) clearly distinguished stem and ground images, indicating that these environments consistently differ with respect to their color composition. The first axis (explaining 78.7% of variation) captured the substantial color variation within each environment. Grasshopper color variation was a subset of environmental variation.

Color differed between banded and uniform grasshoppers (PERMANOVA $R^2 = 0.24$, $P < 0.0001$). Across all collections, the color of uniform morphs was on average more similar to stems than the ground (mean EMD = 0.15 vs 0.22, LMM $P < 2 \times 10^{-16}$, Fig. 6A). By contrast, the color of banded morphs spanned a gradient between "stem-like" and "ground-like" in PCoA space (Fig. 6B). Banded morphs were also more similar to stems than the ground on average, but this difference was slight (mean EMD = 0.17 vs. 0.18, LMM $P = 0.0016$).

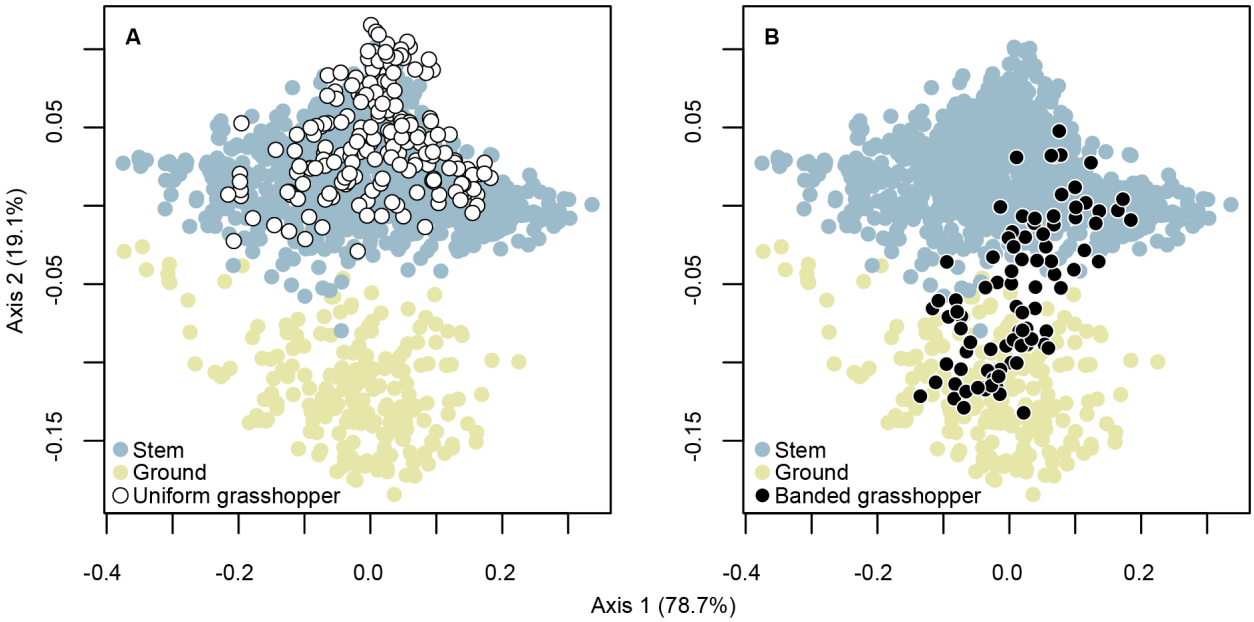

Figure 6. Result of principal coordinates analysis (PCoA) of color variation among all images. **A.** Uniform grasshoppers encompass a similar range of color variation as stems. **B.** Banded grasshoppers span color variation found in both stem and ground environments.

Uniform grasshoppers match local stem color

Within each morph, color differed across sites (banded PERMANOVA $R^2 = 0.71$, $P < 0.0001$; uniform $R^2 = 0.60$, $P < 0.0001$, Fig. 7). Consistent with the hypothesis that predation selects for background matching, uniform grasshoppers more closely resembled local stems than they did stems from other sites (0.141 vs 0.164, LMM $P < 1.2 \times 10^{-13}$). Surprisingly, uniform grasshoppers were also less similar in color to the local ground than they were to the ground at other sites (0.220 vs. 0.170, LMM $P < 2.2 \times 10^{-16}$). By contrast, banded grasshoppers were no more similar in color to local stems ($P = 0.16$) or ground ($P = 0.11$) than they were to these environments at different sites.

Sequencing performance

We recovered a median of 454,323 reads passing QC for each individual (s.d. 388,268, range = 34,266 – 3,549,746). However, the utility of these data was hampered by PCR duplicates, off-target reads, and strong batch effects (Fig. 9). The average rate of PCR duplicates ranged from 0.42 – 0.58 among six library pools, and proportion on-target reads (that is, reads mapping to our reference set of RAD loci) ranged from 0.125 – 0.21). The median number of samples assembled for each individual was 5,000, but of these, a median of 430 loci per sample met our filtering criteria. Notwithstanding these shortcomings, we used the resulting genotype data to test for an association with color pattern.

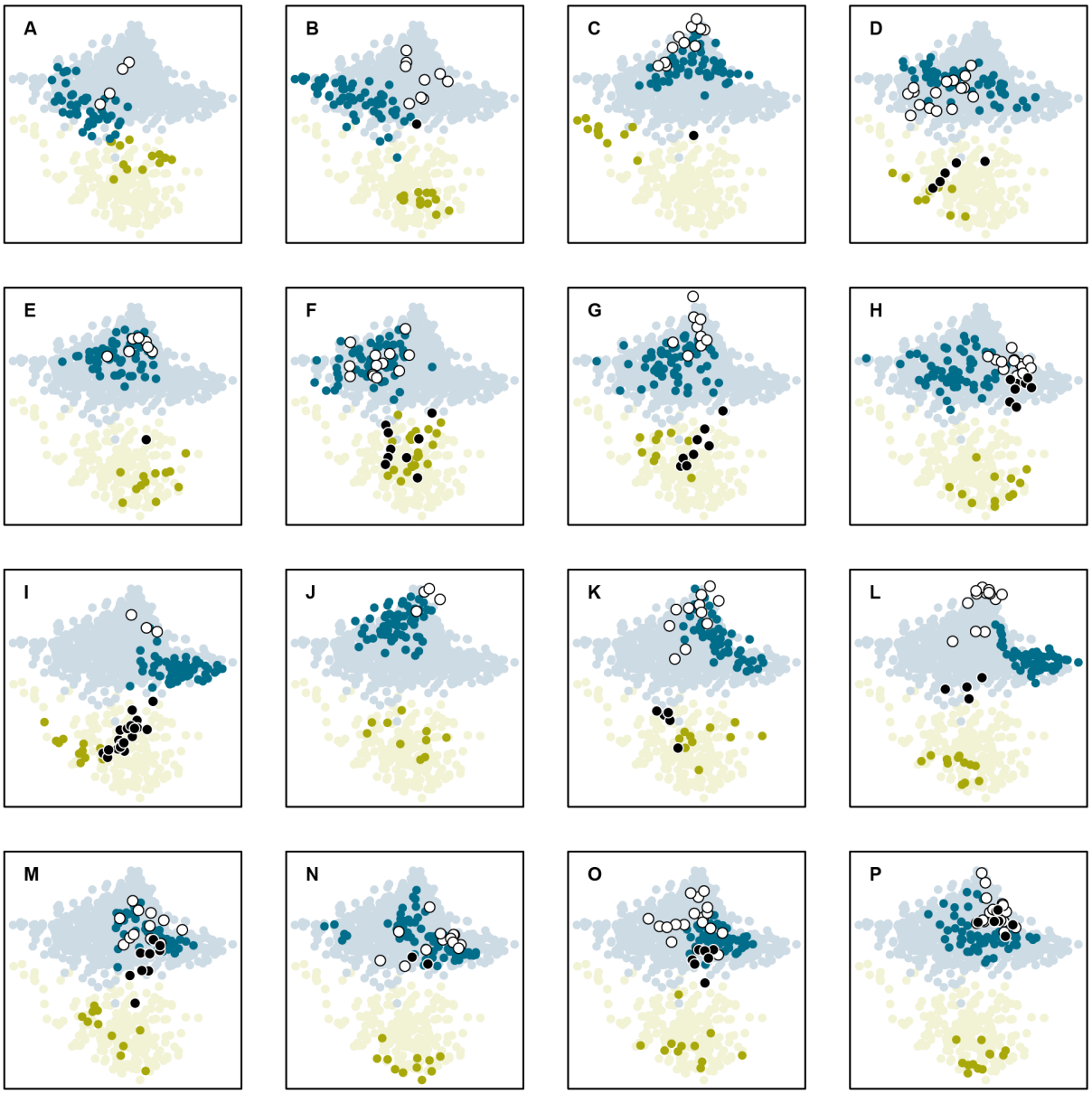

Figure 7. Comparison of color variation within sites. Darkened blue and yellow points show PCoA scores for local stem and ground images, respectively. White and black points show scores for uniform and banded grasshoppers, respectively. Labels correspond to sites in Fig. 1.

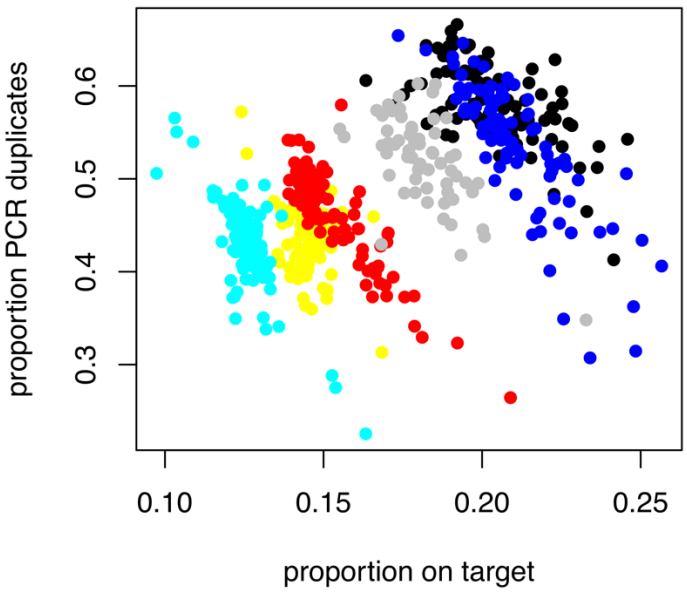

Figure 8. Summary of sequencing performance, showing proportion of PCR duplicates and proportion of reads on-target. Points are colored by the identity of sample pools. These data demonstrate strong batch effects resulting in uneven data quality across pools.

A single RAD locus is strongly associated with color pattern

Among 2,447 loci in our assembly, we identified a single locus that was strongly associated with color morph (Fisher's exact test, $P < 1 \times 10^{-27}$, Fig. 9).

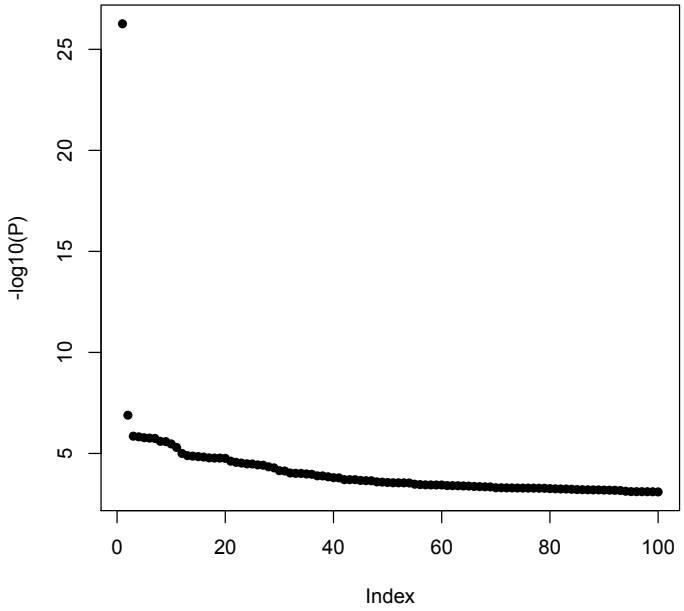

Figure 9. Results of Fisher's exact tests for the top 100 strongest associations between RAD genotype and pattern morph.

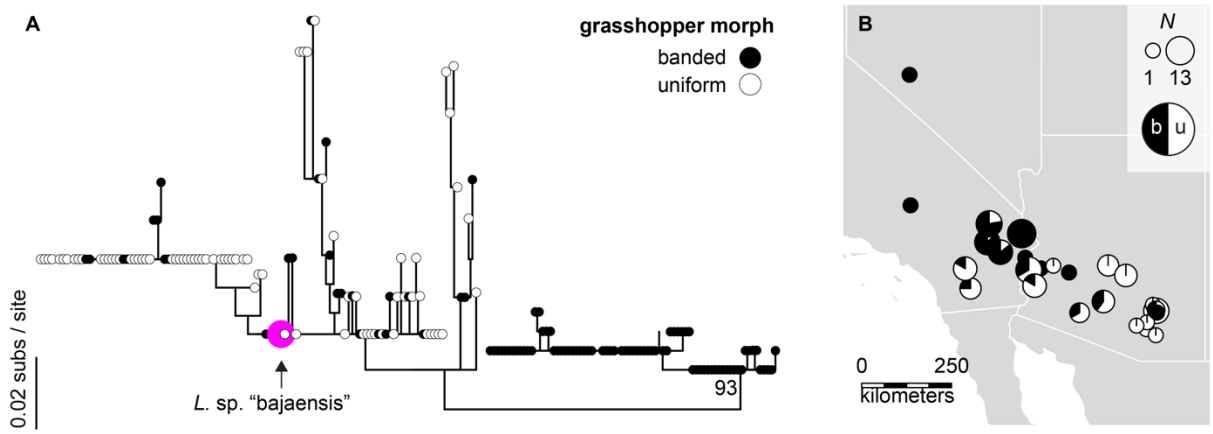

Figure 10. A. Genealogy of haplotypes at a pattern-associated locus. Tip labels indicate pattern phenotype (black = banded, white = uniform). A haplotype from a sister species, *L.* sp. "bajaensis" is shown in magenta. Genealogy is midpoint-rooted. Bootstrap support for the split between banded and uniform alleles is reported beneath the banded clade. **B.** The geographic distribution of samples represented in panel B (*L.* sp. "bajaensis" not shown). Size of circles is proportional to number of samples. Pie wedges indicate proportion banded (black) and uniform (white) individuals.

The genealogy of the pattern-associated locus revealed two striking patterns (Fig. 10). First, two distinct allelic classes segregate within populations. One allele was found only in banded individuals ("banded allele"), and six heterozygotes at the pattern-associated locus were all banded. This suggests the banded allele is dominant to the uniform allele. The six heterozygotes included three males and three females, demonstrating 1) allelic dominance is the same in both sexes, and 2) the pattern-associated locus must be autosomal, since grasshoppers have an XX / XO sex determination system.

Second, a haplotype from the sister species to *L. coquilletti* (*L.* sp. "bajaensis"), is nested within a clade of uniform alleles from *L. coquilletti*. This, along with phenotypic evidence that *L.* sp. "bajaensis" have two pattern morphs similar to *L. coquilletti* (Fig 2), suggests that the pattern-associated locus harbors trans-species polymorphism.

However, analysis of Tajima's *D* in three populations did not support the hypothesis of balancing selection at the pattern associated locus (Table 5).

Table 5. Calculations of Tajima's *D* and tests for deviation from 0.

site	*N* haplotypes	*D*	*P*
C	11	-0.127	0.939
D	19	-1.671	0.078
L	14	-0.438	0.712

DISCUSSION

Sexually antagonistic selection (SA) is common in throughout genomes (Mank, 2017) and across species (Cox & Calsbeek, 2009). Amidst a resurgent research program on balancing selection (Charlesworth, 2006; Fijarczyk & Babik, 2015; Gao et al., 2015; Llaurens et al., 2017), SA remains understudied as process promoting genetic diversity. Understanding how – and for how long – such polymorphisms are maintained is essential to explain the apparent excess of genetic variance in fitness due to balancing selection (Charlesworth, 2015). To do so, studies must characterize antagonistic phenotypes, identify the environmental context of selection, and document the history of loci associated with sexual antagonism. Combining these approaches, we found that unresolved SA results in long-term maintenance of a crypsis polymorphism in the desert clicker grasshopper.

Crypsis is sexually antagonistic

Quantitative comparisons of color and pattern established that crypsis phenotypes are correlated between sexes, and phenotype-environment correlations revealed that different cryptic phenotypes are suited to each environment. While uniform morphs match local stems, the color of banded morphs often resembled that of the ground, and the frequency of banded morphs covaried with the strength of ground patterning. Because males are only found on stems while females are especially exposed to predation on the ground, this suggests that selection for crypsis is sexually antagonistic.

Our finding that crypsis is differentially selected between sexes mirrors studies in adders (Forsman, 1995) and pygmy grasshoppers (Forsman, 2018), where life history differences favor alternative morphs in each sex. In these species, selection for dark coloration to aid thermoregulation trades off with selection for more cryptic phenotypes that minimize predation. Variation in the strength of antagonism can result in geographic variation in morph frequencies, as we observed for *L. coquilletti*.

We next sought to understand the history of SA in *L. coquilletti* through the evolution of loci associated with cryptic phenotypes. We found that pattern morph was strongly associated with a biallelic, autosomal locus with equivalent dominance in each sex. Trans-species polymorphism at this locus indicated that balancing selection has persisted since before the divergence of *L. coquilletti* and *L.* sp. "bajaensis". In sum, our phenotypic, ecological, and genetic data suggests that unresolved sexual antagonism has maintained multiple crypsis morphs since before the divergence of the desert and baja clickers (perhaps millions of years).

Possible forms of selection on females

Several selective processes may underlie the approximately linear relationship we observed between ground pattern and banded morph frequency. If the banded morph is directionally selected in females, the selective advantage of banded morphs may increase with the strength of ground patterning and thus determine equilibrium morph frequencies in a population. Nearly all theory regarding balancing selection via SA has assumed such opposing directional selection between sexes (c.f. Mokkonen et al. 2011). In light of the equal dominance and autosomal inheritance of the pattern-associated locus in *L. coquilletti*, this implies that selection would have to be relatively strong to maintain a stable polymorphism (Kidwell et al., 1977; Patten & Haig, 2009).

Alternatively, selection on females may itself maintain both morphs via spatially varying or negative frequency-dependent selection. Because females are subject to predation on both stems

and the ground, female fitness may be maximized at intermediate morph frequencies. The distribution of phenotypes in a population would thus reflect selective tension within and between sexes. Such tension has been identified in the European serin (*Serinus serin*), for which bill shape is under stabilizing selection in males but directional selection females. Both female morphs may also be maintained by apostatic selection, a common form of negative frequency-dependence that arises when predators preferentially consume common prey types (Bond, 2007). Mokkonen and colleagues (2011) have shown that the interaction of SA and negative frequency-dependence results in more stable polymorphisms. Such a synergy between SA and an additional form of balancing selection may permit the long-term maintenance of banding polymorphism in the desert clicker under a broad range of selection coefficients.

Long-term sexual antagonism without resolution

Our finding of trans-species polymorphism shows that the alleles associated with antagonistic phenotypes are ancient. It is therefore remarkable that SA has not been partially or completely resolved through dominance reversals, sex-linked modifiers, or other mechanisms (Lande, 1980; Bonduriansky & Chenoweth, 2009). One potential explanation depends again upon the form of selection affecting females. If each morph is favored in females under some conditions, selection should preserve the capacity of female to express both. This implies that an interaction of SA and other forms of balancing selection may result in especially persistent antagonistic phenotypes.

Remaining challenges: trans-species polymorphism

Several tasks remain to more robustly characterize trans-species polymorphism. We will address them as follows. First, we will infer a time-calibrated phylogeny of *Ligurotettix* species to place a lower bound on the age of the balanced polymorphism. Second, we will more extensively sequence haplotypes from *L.* sp. "bajaensis" to determine whether this species shares the banded allele in addition to the uniform allele.

Absence of molecular evidence for balancing selection

The absence of evidence for balancing selection from Tajima's *D* may be due to historical factors or methodological biases. Expansion from a population bottleneck can depress values of *D* genome-wide. *L. coquilletti* likely experienced such a recent expansion following glacial retreat and the expansion of North American deserts over the past 12,000 years. Therefore, a more appropriate test for balancing selection would be to compare the value of Tajima's *D* at the pattern-associated locus to the genome-wide distribution. This work is ongoing. In addition, due to challenges with DNA sequencing, the haplotypes we recovered do not approximate a random draw from their source populations. This bias may also result in inaccurate estimates of *D*. Further sequencing at pattern-associated loci will alleviate this issue.

Future directions

A number of questions remain outstanding in this system. For example, it remains unclear what features of banded grasshoppers may enhance crypsis on the ground. Options include background matching, where the color and pattern of an organism resembles a random sample of the environment (Endler, 1978), or disruptive patterning, where high-contrast patterns create false body lines and impede predator detection (Stevens & Merilaita, 2009). These possibilities can now be distinguished with analyses of full-spectral images and models of predator vision (van den Berg et al., 2019) even when field experiments (e.g., Cuthill et al. 2005) are infeasible.

The causative genetic variants that determine color morph remain obscure. However, a strong candidate gene is *WntA*, which has been implicated in the patterning of melanin in butterflies (Martin et al., 2012; Mazo-Vargas et al., 2017; Lewis et al., 2019). Future candidate-gene and unbiased scans for association may illuminate genes responsible for crypsis variation.

REFERENCES

Abrahamson W.G. & Weis A.E. (1997) *Evolutionary Ecology Across Three Trophic Levels: Goldenrods, Gallmakers, and Natural Enemies.* Princeton University Press,

Abràmoff M.D., Magalhães P.J., & Ram S.J. (2004) Image processing with imageJ. *Biophotonics International*, **11**, 36–41.

Agrawal A.F. & Whitlock M.C. (2012) Mutation load: the fitness of individuals in populations where deleterious alleles are abundant. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 115–135.

Aiello-Lammens M.E., Boria R.A., Radosavljevic A., Vilela B., & Anderson R.P. (2015) spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, **38**, 541–545.

Ali O.A., O'Rourke S.M., Amish S.J., Meek M.H., Luikart G., Jeffres C., & Miller M.R. (2016) Rad capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics*, **202**, 389–400.

Allison A.C. (1954) Protection afforded by sickle-cell trait against subtertian malarial infection. *British Medical Journal*, **1**, 290–294.

Anderson M.J. (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, **26**, 32–46.

Anderson R.P., Peterson A.T., & Gómez-Laverde M. (2002) Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos*, **98**, 3–16.

Anderson R.P. (2017) When and how should biotic interactions be considered in models of species niches and distributions? *Journal of Biogeography*, **44**, 8–17.

Arvanitis L., Wiklund C., & Ehrlén J. (2007) Butterfly seed predation: effects of landscape characteristics, plant ploidy level and population structure. *Oecologia*, **152**, 275–285.

Arvanitis L., Wiklund C., Münzbergova Z., Dahlgren J.P., & Ehrlén J. (2010) Novel antagonistic interactions associated with plant polyploidization influence trait selection and habitat preference. *Ecology Letters*, **13**, 330–337.

Austin M. (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1–19.

Avise J.C. (1998) The history and purview of phylogeography: a personal reflection. *Molecular Ecology*, **7**, 371–379.

Bakker E.G., Toomajian C., Kreitman M., & Bergelson J. (2006) A genome-wide survey of R gene polymorphisms in *Arabidopsis*. *The Plant Cell*, **18**, 1803–1818.

Bakovic V., Schuler H., Schebeck M., Feder J.L., Stauffer C., & Ragland G.J. (2019) Host plant-related genomic differentiation in the European cherry fruit fly, *Rhagoletis cerasi* (L., 1758) (Diptera: Tephritidae). *Molecular Ecology*, 4648–4666.

Bao W., Kojima K.K., & Kohany O. (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, **6**, 4–9.

Barbour M.G. (1968) Germination requirements of the desert shrub *Larrea divaricata*. *Ecology*, **49**, 915–923.

Barbour M. (1969) Patterns of genetic similarity between *Larrea divaricata* of North and South America. *American Midland Naturalist*, **81**, 54–67.

Barker B.S., Andonian K., Swope S.M., Luster D.G., & Dlugosch K.M. (2017) Population genomic analyses reveal a history of range expansion and trait evolution across the native

and invaded range of yellow starthistle (*Centaurea solstitialis*). *Molecular Ecology*, **26**, 1131–1147.
Barker M.S., Arrigo N., Baniaga A.E., Li Z., & Levin D.A. (2015) On the relative abundance of autopolyploids and allopolyploids. *New Phytologist*, **210**, 391–398.
Barker M.S., Husband B.C., & Pires J.C. (2016) Spreading Winge and flying high: The evolutionary importance of polyploidy after a century of study. *American Journal of Botany*, **103**, 1139–1145.
Barratt C.D., Bwong B.A., Jehle R., Liedtke H.C., Nagel P., Onstein R.E., Portik D.M., Streicher J.W., & Loader S.P. (2018) Vanishing refuge? Testing the forest refuge hypothesis in coastal East Africa using genome-wide sequence data for seven amphibians. *Molecular Ecology*, **27**, 4289–4308.
Barson N.J., Aykanat T., Hindar K., Baranski M., Bolstad G.H., Fiske P., Jacq C., Jensen A.J., Johnston S.E., Karlsson S., Kent M., Moen T., Niemelä E., Nome T., Næsje T.F., Orell P., Romakkaniemi A., Sægrov H., Urdal K., Erkinaro J., Lien S., & Primmer C.R. (2015) Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, **528**, 405–408.
Bates D., Maechler M., Bolker B., & Walker S. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48.
Bayona-Vásquez N.J., Glenn T.C., Kieran T.J., Finger J.W., Louha S., Troendle N., Diaz-Jaimes P., Mauricio R., & Faircloth B.C. (2019) Adapterama III : Quadruple-indexed , double / triple-enzyme RADseq libraries. *PeerJ*, 1–25.
van Belleghem S.M., Papa R., Ortiz-Zuazaga H., Hendrickx F., Jiggins C.D., Owen McMillan W., & Counterman B.A. (2018) patternize: An R package for quantifying colour pattern variation. *Methods in Ecology and Evolution*, **9**, 390–398.
van den Berg C.P., Troscianko J., Endler J.A., Marshall N.J., & Cheney K.L. (2019) Quantitative Colour Pattern Analysis (QCPA): a comprehensive framework for the analysis of colour patterns in nature. *bioRxiv*, 1–37.
Bergelson J., Kreitman M., Stahl E.A., & Tian D. (2001) Evolutionary dynamics of plant *R*-genes. *Science*, **292**, 2281 LP – 2285.
Bergland A.O., Behrman E.L., O'Brien K.R., Schmidt P.S., & Petrov D.A. (2014) Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila. *PLoS Genetics*, **10**, .
Bi K., Vanderpool D., Singhal S., Linderoth T., Moritz C., & Good J.M. (2012) Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics*, **13**, 403.
Bohnstedt C.F. & Mabry T.J. (1979) The volatile constituents of the genus *Larrea* (Zygophyllaceae). *Revista Latinoamericana de Química*, **10**, 128–131.
Bohonak A.J.. (1994) Dispersal, gene flow, and population structure. *The Quarterly Review of Biology*, **74**, 21–45.
Bond A.B. (2007) The evolution of color polymorphism: crypticity, searching images, and apostatic selection. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 489–514.
Bonduriansky R. & Chenoweth S.F. (2009) Intralocus sexual conflict. *Trends in Ecology and Evolution*, **24**, 280–288.
Borcard D. & Legendre P. (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.
Borcard D., Legendre P., & Drapeau P. (1992) Partialling out the spatial component of

ecological variation. *Ecology*, **73**, 1045–1055.
Bradburd G.S., Ralph P.L., & Coop G.M. (2013) Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, **67**, 3258–3273.
Bray J.R. & Curtis J.T. (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, **27**, 325–349.
Bruner L. (1889) New North American Acrididae, found north of the Mexican boundary. *Proceedings of the United States National Museum*, **7**, 47–83.
Cole K.L. (1986) The Lower Colorado River Valley: a Pleistocene desert. *Quaternary Research*, **25**, 392–400.
Coyne J.A. & Orr H.A. (2004) *Speciation.* Sinauer Associates, Sunderland, MA.
Chapman R., Bernays E., & Wyatt T. (1988) Chemical aspects of host-plant specificity in three *Larrea*-feeding grasshoppers. *Journal of Chemical Ecology*, **14**, 561–579.
Charlesworth B. (2015) Causes of natural variation in fitness: evidence from studies of Drosophila populations. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 1662–1669.
Charlesworth D. (2006) Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, **2**, 379–384.
Clark A.G. (1997) Neutral behavior of shared polymorphism. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 7730–7734.
Connallon T. & Chenoweth S.F. (2019) Dominance reversals and the maintenance of genetic variation for fitness. *PLoS Biology*, **17**, 1–11.
Connallon T. & Clark A.G. (2012) A general population genetic framework for antagonistic selection that accounts for demography and recurrent mutation. *Genetics*, **190**, 1477–1489.
Connallon T. & Clark A.G. (2014a) Balancing selection in species with separate sexes: insights from Fisher's geometric model. *Genetics*, **197**, 991–1006.
Connallon T. & Clark A.G. (2014b) Evolutionary inevitability of sexual antagonism. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20132123.
Connallon T., Cox R.M., & Calsbeek R. (2010) Fitness consequences of sex-specific selection. *Evolution*, **64**, 1671–1682.
Cox R.M. & Calsbeek R. (2009) Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *American Naturalist*, **173**, 176–187.
Curtsinger J.W. (1980) On the opportunity for polymorphism with sex-linkage or haplodiploidy. *Genetics*, **96**, 995–1006.
Cuthill I.C., Stevens M., Sheppard J., Maddocks T., Parraga C.A., & Troscianko T.S. (2005) Disruptive coloration and background pattern matching. *Nature*, **434**, 72–74.
Dormann C.F., Elith J., Bacher S., Buchmann C., Carl G., Carré G., Marquéz J.R.G., Gruber B., Lafourcade B., Leitão P.J., Münkemüller T., Mcclean C., Osborne P.E., Reineking B., Schröder B., Skidmore A.K., Zurell D., & Lautenbach S. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 027–046.
Delph L.F. & Kelly J.K. (2014) On the importance of balancing selection in plants. *New Phytologist*, **201**, 45–56.
Dempster E.R. (1956) Maintenance of genetic heterogeneity. *Cold Spring Harbor Symposia in Quantitative Biology*, **20**, 25:32.
van Diepen L.T.A., Olson A., Ihrmark K., Stenlid J., & James T.Y. (2013) Extensive trans-specific polymorphism at the mating type locus of the root decay fungus heterobasidion.

Molecular Biology and Evolution, **30**, 2286–2301.
van Doorn G.S. (2009) Intralocus sexual conflict. *Annals of the New York Academy of Sciences*, **1168**, 52–71.
Dray S., Bauman D., Blanchet G., Borcard D., Clappe S., Guenard G., Jombart T., Larocque G., Legendre P., Madi N., & Wagner H.H. (2019) adespatial: Multivariate Multiscale Spatial Analysis. .
Drès M. & Mallet J. (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **357**, 471–92.
Driscoe A.L., Nice C.C., Busbee R.W., Hood G.R., Egan S.P., & Ott J.R. (2019) Host plant associations and geography interact to shape diversification in a specialist insect herbivore. *Molecular Ecology*, **28**, 4197–4211.
Dudík M., Phillips S.J., & Schapire R.E. (2005) Correcting sample selection bias in maximum entropy density estimation. *Advances in Neural Information Processing Systems* (ed. by Y. Weiss, B. Schölkopf, and J.C. Platt), pp. 323–330. MIT Press, Cambridge, MA.
Eaton D.A.R. (2014) PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.
Edgar R.C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.
Endler J.A. (1978) A predator's view of animal color patterns. *Evolutionary Biology*, **11**, 320–364.
Farrell B. & Mitter C. (1990) Phylogenesis of insect plant interactions: have *Phyllobrotica* leaf beetles (Chrysomelidae) and the Lamiales diversified in parallel? *Evolution*, **44**, 1389–1403.
Farrell B.D. (2001) Evolutionary assembly of the milkweed fauna: cytochrome oxidase I and the age of *Tetraopes* beetles. *Molecular Phylogenetics and Evolution*, **18**, 467–478.
Feder J.L., Chilcote C.A., & Bush G.L. (1988) Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature*, **366**, 61–64.
Fick S.E. & Hijmans R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302–4315.
Figueroa F., Günther E., & Klein J. (1988) MHC polymorphism pre-dating speciation. *Nature*, **335**, 265–267.
Fijarczyk A. & Babik W. (2015) Detecting balancing selection in genomes: limits and prospects. *Molecular Ecology*, **24**, 3529–3545.
Fisher R.A. (1922) On the dominance ratio. *Proceedings of the Royal Society of Edinburgh*, **42**, 321–341.
Fitzpatrick M.C. & Hargrove W.W. (2009) The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation*, **18**, 2255–2261.
Foerster K., Coulson T., Sheldon B.C., Pemberton J.M., Clutton-Brock T.H., & Kruuk L.E.B. (2007) Sexually antagonistic genetic variation for fitness in red deer. *Nature*, **447**, 1107–1110.
Forbes A.A., Devine S.N., Hippee A.C., Tvedte E.S., Ward A.K.G., Widmayer H.A., & Wilson C.J. (2017) Revisiting the particular role of host shifts in initiating insect speciation. *Evolution*, **71**, 1126–1137.
Forsman A. (1995) Opposing fitness consequences of colour pattern in male and female snakes. *Journal of Evolutionary Biology*, **8**, 53–70.
Forsman A. (2018) On the role of sex differences for evolution in heterogeneous and changing

fitness landscapes: insights from pygmy grasshoppers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **373**, 20170429.
Fox J. & Weisberg S. (2019) *An R Companion to Applied Regression.* Sage, Thousand Oaks, CA.
Fry J.D. (2010) The genomic location of sexually antagonistic variation: Some cautionary comments. *Evolution*, **64**, 1510–1516.
Funk D.J., Egan S.P., & Nosil P. (2011) Isolation by adaptation in *Neochlamisus* leaf beetles: host-related selection promotes neutral genomic divergence. *Molecular Ecology*, 4671–4682.
Gagné R. & Waring G. (1990) The *Asphondylia* (Cecidomyiidae: Diptera) of creosote bush (*Larrea tridentata*) in North America. *Proceedings of the Entomological Society of Washington*, **92**, 649–671.
Gagné R.J. (1989) *The Plant-Feeding Gall Midges of North America.* Cornell University Press, Ithaca, NY.
Gagné R.J. & Jaschof M. (2014) *A Catalog of the Cecidomyiidae (Diptera) of the World. 3rd Edition. Digital Version 2*. Available from https://www.ars.usda.gov/ARSUserFiles/80420580/Gagne_2014_World_Cecidomyiidae_Catalog_3rd_Edition.pdf.
Gao Z., Przeworski M., & Sella G. (2015) Footprints of ancient-balanced polymorphisms in genetic variation data from closely related species. *Evolution*, **69**, 431–446.
Gillespie J.H. (1978) A general model to account for enzyme variation in natural populations. V. The SAS-CFF model. *Theoretical Population Biology*, **14**, 1–45.
Glennon K.L., Ritchie M.E., & Segraves K.A. (2014) Evidence for shared broad-scale climatic niches of diploid and polyploid plants. *Ecology Letters*, **17**, 574–582.
Gompert Z., Comeault A. a, Farkas T.E., Feder J.L., Parchman T.L., Buerkle C.A., & Nosil P. (2014) Experimental evidence for ecological selection on genome variation in the wild. *Ecology Letters*, **17**, 369–79.
Gregory T.R. (2019) *Animal Genome Size Database.* http://www.genomesize.com
Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., & Gascuel O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.
Hafner D.J. & Riddle B.R. (2011) Barriers and Boundaries of North American Warm Deserts: An Evolutionary Perspective. *Palaeogeography and Palaeobiogeography: Biodiversity in Space and Time* pp. 75–114.
Haldane J.B.S. (1957) The cost of natural selection. *Journal of Genetics*, **33**, 511–524.
Haldane J.B.S. (1962) Conditions for a stable polymorphism at an autosomal locus. *Nature*, **193**, 1108.
Hall D.R., Amarawardana L., Cross J. V., Francke W., Boddum T., & Hillbur Y. (2012) The chemical ecology of Cecidomyiid midges (Diptera: Cecidomyiidae). *Journal of Chemical Ecology*, **38**, 2–22.
Halverson K., Heard S.B., Nason J.D., & Stireman J.O. (2008a) Differential attack on diploid, tetraploid, and hexaploid *Solidago altissima* L. by five insect gallmakers. *Oecologia*, **154**, 755–761.
Halverson K., Heard S.B., Nason J.D., & Stireman J.O. (2008b) Origins, distribution, and local co-occurrence of polyploid cytotypes in *Solidago altissima* (Asteraceae). *American Journal of Botany*, **95**, 50–58.

Harris M.O., Stuart J.J., Mohan M., Nair S., Lamb R.J., & Rohfritsch O. (2003) Grasses and gall midges: plant defense and insect adaptation. *Annual Review of Entomology*, **48**, 549–577.
Hawkins B.A., Goeden R.D., & Gagné R.J. (1986) Ecology and taxonomy of the *Asphondylia* spp. (Diptera: Cecidomyiidae) forming galls on *Atriplex* spp. (Chenopodiaceae) in southern California. *Entomography*, **4**, 55–107.
Hedrick P.W. (2012) What is the evidence for heterozygote advantage selection? *Trends in Ecology and Evolution*, **27**, 698–704.
Highland H.A. (1964) Life history of *Asphondylia ilicicola* (Diptera: Cecidomyiidae), a pest of American holly. *Journal of Economic Entomology*, **57**, 81–83.
Hijmans R.J. (2016) raster: Geographic data analysis and modeling. R package version 2.5-8. Available from https://CRAN.R-project.org/package=raster.
Hijmans R.J., Phillips S., Leathwick J., & Elith J. (2017) dismo: species distribution modeling. R package version 1.1-4. Available from https://CRAN.R-project.org/package=dismo.
Hoffberg S.L., Kieran T.J., Catchen J.M., Devault A., Faircloth B.C., Mauricio R., & Glenn T.C. (2016) RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, **16**, 1264–1278.
Holmgren C. a., Betancourt J.L., Peñalba M.C., Delgadillo J., Zuravnsky K., Hunter K.L., Rylander K.A., & Weiss J.L. (2014) Evidence against a Pleistocene desert refugium in the Lower Colorado River Basin. *Journal of Biogeography*, **41**, 1769–1780.
Hori M. (1993) Frequency-dependent natural selection in the handedness of scale-ating cichlid fish. *Science*, **260**, 216–219.
Huang Y., Wright S.I., & Agrawal A.F. (2014) Genome-wide patterns of genetic variation within and among alternative selective regimes. *PLoS Genetics*, **10**, e1004527.
Huggins T.R. (2008) *Gall morphology and the effects of host plant water status on the* Asphondylia auripila *group on* Larrea tridentata *in the Mojave Desert, Granite Mountains, California* (Unpublished doctoral dissertation). University of California Los Angeles, Los Angeles, CA.
Hunter K.L., Betancourt J.L., Riddle B.R., Van Devender T.R., Cole K.L., & Spaulding W.G. (2001) Ploidy race distributions since the Last Glacial Maximum in the North American desert shrub, *Larrea tridentata*. *Global Ecology and Biogeography*, **10**, 521–533.
Janz N. & Thompson J.N. (2002) Plant polyploidy and host expansion in an insect herbivore. *Oecologia*, **130**, 570–575.
Jombart T. (2008) Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
Jordan C.Y. & Charlesworth D. (2012) The potential for sexually antagonistic polymorphism in different genome regions. *Evolution*, **66**, 505–516.
Joy J.B. & Crespi B.J. (2007) Adaptive radiation of gall-inducing insects within a single host-plant species. *Evolution*, **61**, 784–95.
Kanno H. & Harris M.O. (2000) Physical features of grass leaves influence the placement of eggs within the plant by the Hessian fly. *Entomologia Experimentalis et Applicata*, **96**, 69–80.
Kao R.H. (2008a) Origins and widespread distribution of co-existing polyploids in *Arnica cordifolia* (Asteraceae). *Annals of Botany*, **101**, 145–152.
Kao R.H. (2008b) Implications of polyploidy in the host plant of a dipteran seed parasite. *Western North American Naturalist*, **68**, 225–230.
Karasov T.L., Kniskern J.M., Gao L., Deyoung B.J., Ding J., Dubiella U., Lastra R.O., Nallu S.,

Roux F., Innes R.W., Barrett L.G., Hudson R.R., & Bergelson J. (2014) The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature*, **512**, 436–440.

Kidwell J.F., Clegg M.T., Stewart F.M., & Prout T. (1977) Regions of stable equilibria for models of differential selection in the two sexes under random mating. *Genetics*, **85**, 171–183.

Kramer-Schadt S., Niedballa J., Pilgrim J.D., Schröder B., Lindenborn J., Reinfelder V., Stillfried M., Heckmann I., Scharf A.K., Augeri D.M., Cheyne S.M., Hearn A.J., Ross J., Macdonald D.W., Mathai J., Eaton J., Marshall A.J., Semiadi G., Rustam R., Bernard H., Alfred R., Samejima H., Duckworth J.W., Breitenmoser-Wuersten C., Belant J.L., Hofer H., & Wilting A. (2013) The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions*, **19**, 1366–1379.

Kremen C., Cameron A., Moilanen A., Phillips S.J., Thomas C.D., Beentje H., Dransfield J., Fisher B.L., Glaw F., Good T.C., Harper G.J., Hijmans R.J., Lees D.C., Louis E., Nussbaum R.A., Raxworthy C.J., Razafimpahanana A., Schatz G.E., Vences M., Vieites D.R., Wright P.C., & Zjhra M.L. (2008) Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science*, **320**, 222–226.

Lande R. (1980) Sexual eimorphism, sexual selection, and adaptation in polygenic characters. *Evolution*, **34**, 292.

Laport R. & Minckley R. (2013) Cytogeography of *Larrea tridentata* at the Chihuahuan–Sonoran Desert ecotone. *USDA Forest Service Proceedings*, 218–224.

Laport R.G., Hatem L., Minckley R.L., & Ramsey J. (2013) Ecological niche modeling implicates climatic adaptation, competitive exclusion, and niche conservatism among *Larrea tridentata* cytotypes in North American deserts. *The Journal of the Torrey Botanical Society*, **140**, 349–363.

Laport R.G., Minckley R.L., & Ramsey J. (2012) Phylogeny and cytogeography of the North American creosote bush (*Larrea tridentata*, Zygophyllaceae). *Systematic Botany*, **37**, 153–164.

Laport R.G., Minckley R.L., & Ramsey J. (2016) Ecological distributions, phenological isolation, and genetic structure in sympatric and parapatric populations of the *Larrea tridentata* polyploid complex. *American Journal of Botany*, **103**, 1358–1374.

Laport R.G. & Ng J. (2017) Out of one, many: The biodiversity considerations of polyploidy. *American Journal of Botany*, **104**, 1119–1121.

Laport R.G. & Ramsey J. (2015) Morphometric analysis of the North American creosote bush (*Larrea tridentata*, Zygophyllaceae) and the microspatial distribution of its chromosome races. *Plant Systematics and Evolution*, **301**, 1581–1599.

Leffler E.M., Gao Z., Pfeifer S., Ségurel L., Auton A., Venn O., Bowden R., Bontrop R., Wall J.D., Sella G., Donnelly P., McVean G., & Przeworski M. (2013) Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science*, **339**, 1578–1582.

Levene H. (1953) Genetic equilibrium when more than one ecological niche is availablle. *The American Naturalist*, **87**, 331-333.

Levin D.A. (1975) Minority cytotype exclusion in local plant populations. *Taxon*, **24**, 35–43.

Levin D.A. (1983) Polyploidy and novelty in flowering plants. *The American Naturalist*, **122**, 1–25.

Lewis J.J., Geltman R.C., Pollak P.C., Rondem K.E., van Belleghem S.M., Hubisz M.J., Munn

P.R., Zhang L., Benson C., Mazo-Vargas A., Danko C.G., Counterman B.A., Papa R., & Reed R.D. (2019) Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing pattern mimicry. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, .

Lewis W.H. (1980) Polyploidy in species populations. *Polyploidy: Biological Relevance* (ed. by W.H. Lewis), pp. 103–144. Plenum Press, New York, NY.

Lewontin R.C. & Hubby J.L. (1966) A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudoobscura. *Genetics*, **54**, 595–609.

Linck E. & Battey C.J. (2019) Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, **19**, 639–647.

Lowry D.B., Rockwood R.C., & Willis J.H. (2008) Ecological reproductive isolation of coast and inland races of Mimulus guttatus. *Evolution*, **62**, 2196–2214.

Llaurens V., Whibley A., & Joron M. (2017) Genetic architecture and balancing selection: the life and death of differentiated variants. *Molecular Ecology*, **26**, 2430–2448.

Long T.A.F., Montgomerie R., & Chippindale A.K. (2006) Quantifying the gender load: Can population crosses reveal interlocus sexual conflict? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **361**, 363–374.

Mabry T.J., DiFeo Jr D.R., Sakakibara M., Bohnstedt Jr C.F., & Siegler D. (1977) The natural products chemistry of *Larrea*. *Creosote Bush: Biology and Chemistry of* Larrea *in New World Deserts* (ed. by T.J. Mabry, A.J.H. Hunziker, and D.R. DiFeo Jr), pp. 115–134. Dowden, Hutchinson & Ross Inc, Stroudsburg, PA.

MacMahon J.A. & Wagner F.H. (1985) Mojave, Sonoran and Chihuahuan Deserts of North America. *Ecosystems of the World, Vol. 12A* (ed. by M. Evanri, I. Noy-Meir, and D. Goodall), pp. 105–198. Amsterdam.

Madlung A. (2013) Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity*, **110**, 99–104.

Mandáková T. & Münzbergová Z. (2006) Distribution and ecology of cytotypes of the *Aster amellus* aggregates in the Czech Republic. *Annals of Botany*, **98**, 845–856.

Mank J.E. (2017) Population genetics of sexual conflict in the genomic era. *Nature Reviews Genetics*, **18**, 721–730.

Martin A., Papa R., Nadeau N.J., Hill R.I., Counterman B.A., Halder G., Jiggins C.D., Kronforst M.R., Long A.D., McMillan W.O., & Reed R.D. (2012) Diversification of complex butterfly wing patterns by repeated regulatory evolution of a *Wnt* ligand. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 12632–12637.

Matthews G., Hangartner S., Chapple D.G., & Connallon T. (2019) Quantifying maladaptation during the evolution of sexual dimorphism. *Proceedings of the Royal Society B: Biological Sciences*, **286**, .

Mazo-Vargas A., Concha C., Livraghi L., Massardo D., Wallbank R.W.R., Zhang L., Papador J.D., Martinez-Najera D., Jiggins C.D., Kronforst M.R., Breuker C.J., Reed R.D., Patel N.H., McMillan W.O., & Martin A. (2017) Macroevolutionary shifts of WntA function potentiate butterfly wing-pattern diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, 10701–10706.

McArdle B.H. & Anderson M.J. (2001) Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology*, **82**, 290–297.

McCartney-Melstad E., Mount G.G., & Shaffer H.B. (2016) Exon capture optimization in

amphibians with large genomes. *Molecular Ecology Resources*, **16**, 1084–1094.
McIntyre P.J. (2012) Polyploidy associated with altered and broader ecological niches in the *Claytonia perfoliata* (Portulacaceae) species complex. *American Journal of Botany*, **99**, 655–662.
Miao J., Wu Y.Q., Gong Z.J., He Y.Z., Duan Y., & Jiang Y.L. (2013) Long-distance wind-borne dispersal of *Sitodiplosis mosellana* Géhin (Diptera: Cecidomyiidae) in Northern China. *Journal of Insect Behavior*, **26**, 120–129.
Mokkonen M., Kokko H., Koskela E., Lehtonen J., Mappes T., Martiskainen H., & Mills S.C. (2011) Negative frequency-dependent selection of sexually antagonistic alleles in Myodes glareolus. *Science*, **334**, 972–974.
Münzbergová Z., Skuhrovec J., & Maršík P. (2015) Large differences in the composition of herbivore communities and seed damage in diploid and autotetraploid plant species. *Biological Journal of the Linnean Society*, **115**, 270–287.
Mullon C., Pomiankowski A., & Reuter M. (2012) The effects of selection and genetic drift on the genomic distribution of sexually antagonistic alleles. *Evolution*, **66**, 3743–3753.
Muscarella R., Galante P.J., Soley-Guardia M., Boria R.A., Kass J.M., Uriarte M., & Anderson R.P. (2014) ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, **5**, 1198–1205.
Nason J.D., Heard S.B., & Williams F.R. (2002) Host-associated genetic differentiation in the goldenrod elliptical-gall moth, *Gnorimoschema gallaesolidaginis* (Lepidoptera: Gelechiidae). *Evolution*, **56**, 1475–88.
Ney G. & Schul J. (2017) Low genetic differentiation between populations of an endemic prairie katydid despite habitat loss and fragmentation. *Conservation Genetics*, **18**, 1389–1401.
Nosil P. (2012) *Ecological speciation.* Oxford University Press,
Nosil P., Crespi B.J., & Sandoval C.P. (2002) Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature*, **417**, 440–443.
Nuismer S.L. & Thompson J.N. (2001) Plant polyploidy and non-uniform effects on insect herbivores. *Proceedings of the Royal Society B: Biological Sciences*, **268**, 1937–1940.
O'Connor T.K., Laport R.G., & Whiteman N.K. (2019) Polyploidy in creosote bush (*Larrea tridentata*) shapes the biogeography of specialist herbivores. *Journal of Biogeography*, **46**, 597–610.
Oksanen J., Blanchet F.G., Friendly M., Kindt R., Legendre P., McGlinn D., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Szoecs E., & Wagner H. (2017) vegan: community ecology package. R package version 2.4-4. Available from https://CRAN.R-project.org/package=vegan.
Ohta T. (1992) The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, **23**, 263–286.
Owen A.R.G. (1953) A genetical system admitting of two stable equilibria under natural selection. *Heredity*, **7**, 97–102.
Parisod C., Holderegger R., & Brochmann C. (2010) Evolutionary consequences of autopolyploidy. *New Phytologist*, **186**, 5–17.
Patten M.M. & Haig D. (2009) Maintenance or loss of genetic variation under sexual and parental antagonism at a sex-linked locus. *Evolution*, **63**, 2888–2895.
Pearse D.E., Barson N.J., Nome T., Gao G., Campbell M.A., Abadía-Cardoso A., Anderson E.C., Rundio D.E., Williams T.H., Naish K.A., Moen T., Liu S., Kent M., Minkley D.R.,

Rondeau E.B., Brieuc M.S.O., Sandve S.R., Miller M.R., Cedillo L., Baruch K., Hernandez A.G., Ben-Zvi G., Shem-Tov D., Barad O., Kuzishchin K., Garza J.C., Lindley S.T., Koop B.F., Thorgaard G.H., Palti Y., & Lien S. (2019) Sex-dependent dominance maintains migration supergene in rainbow trout. *Nature Ecology and Evolution*, **3**, 1731–1742.

Pearson R.G., Raxworthy C.J., Nakamura M., & Peterson A.T. (2007) Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *Journal of Biogeography*, **34**, 102–117.

Peccoud J., Ollivier A., Plantegenest M., & Simon J.-C. (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 7495–7500.

Van de Peer Y., Mizrachi E., & Marchal K. (2017) The evolutionary significance of polyploidy. *Nature Reviews Genetics*, **18**, 411–424.

Peterson B.K., Weber J.N., Kay E.H., Fisher H.S., & Hoekstra H.E. (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*, **7**, e37135.

Petitpierre B., Broennimann O., Kueffer C., Daehler C., & Guisan A. (2017) Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions. *Global Ecology and Biogeography*, **26**, 275–287.

Petkova D., Novembre J., & Stephens M. (2016) Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, **48**, 94–100.

Phillips S. (2017) maxnet: fitting 'Maxent' species distribution models with 'glmnet.' R package version 0.1.2. Available from https://CRAN.R-project.org/package=maxnet.

Phillips S.J., Anderson R.P., Dudík M., Schapire R.E., & Blair M.E. (2017) Opening the black box: an open-source release of Maxent. *Ecography*, **40**, 887–893.

Phillips S.J., Dudík M., & Schapire R.E. (2004) A maximum entropy approach to species distribution modeling. *21st International Conference on Machine Learning, Banff, Canada*, 655–662.

Phillipsen I.C., Kirk E.H., Bogan M.T., Mims M.C., Olden J.D., & Lytle D.A. (2015) Dispersal ability and habitat requirements determine landscape-level genetic patterns in desert aquatic insects. *Molecular Ecology*, **24**, 54–69.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Radosavljevic A. & Anderson R.P. (2014) Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, **41**, 629–643.

Ramsey J. (2011) Polyploidy and ecological adaptation in wild yarrow. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 7096–7101.

Ramsey J. & Ramsey T.S. (2014) Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**, 1–76.

Ramsey J. & Schemske D.W. (2002) Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics*, **33**, 589–639.

Rasmussen M.D., Hubisz M.J., Gronau I., & Siepel A. (2014) Genome-wide inference of ancestral recombination graphs. *PLoS Genetics*, **10**, e1004342.

Rhoades D.F. (1977) The antiherbivore chemistry of *Larrea*. *Creosote Bush: Biology and Chemistry of* Larrea *in New World Deserts* (ed. by T.J. Mabry, A.J.H. Hunziker, and D.R.

DiFeo Jr), pp. 135–175. Dowden, Hutchinson & Ross, Stroudsburg, Pennsylvania.
Rhoades D. & Cates R. (1976) Toward a general theory of plant antiherbivore chemistry. *Biochemical Interaction Between Plants and Insects* (ed. by J.W. Wallace and R.L. Mansell), Plenum Press, New York.
Rice W.R. (1984) Sex chromosomes and the evolution of sexual dimorphism. *Evolution*, **38**, 735–742.
Rice W.R. & Chippindale A.K. (2001) Intersexual ontogenetic conflict. *Journal of Evolutionary Biology*, **14**, 685–693.
Richardson M.L. & Hanks L.M. (2011) Differences in spatial distribution, morphology, and communities of herbivorous insects among three cytotypes of *Solidago altissima* (Asteraceae). *American Journal of Botany*, **98**, 1595–1601.
Rohland N. & Reich D. (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, **22**, 939–946.
Ronce O. (2014) Geographic Variation, Population Structure, and Migration. *The Princeton Guide to Evolution* (ed. by J.B. Losos, D.A. Baum, D.J. Futuyma, H.E. Hoekstra, R.E. Lenski, and A.J. Moore), pp. 321–327. Princeton University Press, Princeton, NJ.
Rubner Y., Tomasi C., & Guibas L.J. (2000) Earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, **40**, 99–121.
Sanderson S.C. & Stutz H.C. (1994) High chromosome numbers in Mojavean and Sonoran Desert *Atriplex canescens* (Chenopodiaceae). *American Journal of Botany*, **81**, 1045–1053.
Satler J.D. & Carstens B.C. (2016) Phylogeographic concordance factors quantify phylogeographic congruence among co-distributed species in the *Sarracenia alata* pitcher plant system. *Evolution*, **70**, 1105–1119.
Satler J.D. & Carstens B.C. (2017) Do ecological communities disperse across biogeographic barriers as a unit? *Molecular Ecology*, **26**, 3533–3545.
Satler J.D., Zellmer A.J., & Carstens B.C. (2016) Biogeographic barriers drive co-diversification within associated eukaryotes of the *Sarracenia alata* pitcher plant system. *PeerJ*, **4**, e1576.
Schowalter T.D., Lightfoot D.C., & Whitford W.G. (1999) Diversity of arthropod responses to host-plant water stress in a desert ecosystem in southern New Mexico. *The American Midland Naturalist*, **142**, 281–290.
Scudder S.H. (1890) Some genera of Oedipodidae rescued from the Tryxalidae. *Psyche*, **5**, 431–442.
Segraves K.A. (2017) The effects of genome duplications in a community context. *New Phytologist*, **215**, 57–69.
Segraves K.A. & Anneberg T.J. (2016) Species interactions and plant polyploidy. *American Journal of Botany*, **103**, 1326–1335.
Segraves K.A., Thompson J.N., Soltis P.S., & Soltis D.E. (1999) Multiple origins of polyploidy and the geographic structure of *Heuchera grossulariifolia*. *Molecular Ecology*, **8**, 253–262.
Ségurel L., Thompson E.E., Flutre T., Lovstad J., Venkat A., Susan W., Moyse J., Ross S., Gamble K., Sella G., Ober C., & Przeworski M. (2012) The ABO blood group is a trans-species polymorphism in primates. *Proceedings of the National Academy of Sciences*, **109**, 18493–18498.
Sexton J.P., Mcintyre P.J., Angert A.L., & Rice K.J. (2009) Evolution and ecology of species range limits. *Annual Review of Ecology & Systematics*, **40**, 415–436.
Sexton J.P., Hangartner S.B., & Hoffmann A.A. (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*, **68**, 1–15.

Shafer A.B.A. & Wolf J.B.W. (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, **16**, 940–950.
Simon C., Frati F., Beckenbach A., Crespi B., Liu H., & Flook P. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.
Sinervo B. & Lively C.M. (1996) The rock-paper-scissors game and the evolution of alternative male strategies. *Nature*, **380**, 240–243.
Slatkin M. (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.
Smit, A.F.A., Hubley, R. & Green P. *RepeatMasker Open-4.0. 2013-2015.*
Smith C.I., Godsoe W.K.W., Tank S., Yoder J.B., & Pellmyr O. (2008) Distinguishing coevolution from covicariance in an obligate pollination mutualism: asynchronous divergence in Joshua tree and its pollinators. *Evolution*, **62**, 2676–2687.
Smith C.I., Tank S., Godsoe W., Levenick J., Strand E., Esque T., & Pellmyr O. (2011) Comparative phylogeography of a coevolved community: concerted population expansions in Joshua trees and four Yucca moths. *PLoS ONE*, **6**, .
Soltis D.E., Soltis P.S., Schemske D.W., Hancock J.F., John N., Husband B.C., Judd W.S., Hancock J.F., Soltis D.E., Soltis P.S., Schemske D.W., Thompson J.N., Husband B.C., & Judd W.S. (2007) Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon*, **56**, 13–30.
Städler T. & Delph L.F. (2002) Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 11730–11735.
Stamatakis A. (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.
Stebbins G.L. (1949) The evolutionary significance of natural and artificial polyploids in the family Graminae. *Hereditas*, **35**, 461–485.
Stevens M. & Merilaita S. (2009) Defining disruptive coloration and distinguishing its functions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 481–488.
Stireman J.O., Nason J.D., & Heard S.B. (2005) Host-associated genetic differentiation in phytophagous insects: general phenomenon or isolated exceptions? Evidence from a goldenrod-insect community. *Evolution*, **59**, 2573–2587.
Stoddard M.C. & Stevens M. (2010) Pattern mimicry of host eggs by the common cuckoo, as seen through a bird's eye. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 1387–1393.
Surridge A.K. & Mundy N.I. (2002) Trans-specific evolution of opsin alleles and the maintenance of trichromatic colour vision in Callitrichine primates. *Molecular Ecology*, **11**, 2157–2169.
Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **595**, 585–595.
Thompson J.N., Nuismer S.L., & Merg K. (2004) Plant polyploidy and the evolutionary ecology of plant/animal interactions. *Biological Journal of the Linnean Society*, **82**, 511–519.
Thompson J.N., Segraves K.A., Cunningham B.M., Althoff D.M., & Wagner D. (1997) Plant polyploidy and insect/plant interactions. **150**, 730–743.
Thompson K.A., Husband B.C., & Maherali H. (2014) Climatic niche differences between

diploid and tetraploid cytotypes of *Chamerion angustifolium* (Onagraceae). *American Journal of Botany*, **101**, 1868–1875.
Tinkham E.R. (1938) Western Orthoptera attracted to lights. *Journal of the New York Entomological Society*, **46**, 339–353.
Tokuda M. (2012) Biology of Asphondyliini (Diptera: Cecidomyiidae). *Entomological Science*, **15**, 361–383.
Troscianko J. & Stevens M. (2015) Image calibration and analysis toolbox – a free software suite for objectively measuring reflectance, colour and pattern. *Methods in Ecology and Evolution*, **6**, 1320–1331.
Uechi N., Yukawa J., & Yamaguchi D. (2004) Host alternation by gall midges of the genus *Asphondylia* (Diptera: Cecidomyiidae). *Bishop Museum Bulletin in Entomology*, **12**, 53–66.
Uyenoyama M.K. (1997) Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. *Genetics*, **147**, 1389–1400.
Venables W.N. & Ripley B.D. (2002) *Modern Applied Statistics with S.* Springer, New York.
Via S. (1999) Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution*, **53**, 1446–1457.
Vicoso B. & Charlesworth B. (2006) Evolution on the X chromosome: Unusual patterns and processes. *Nature Reviews Genetics*, **7**, 645–653.
Vidal M.C., Quinn T.W., Stireman J.O., Tinghitella R.M., & Murphy S.M. (2019) Geography is more important than host plant use for the population genetic structure of a generalist insect herbivore. *Molecular Ecology*, **28**, 4317–4334.
Wang G. & Greenfield M. (1994) Ontogeny of territoriality in the desert clicker *Ligurotettix coquilletti* (Orthoptera: Acrididae). *Journal of Insect Behavior*, **7**, 327–342.
Wang I.J. & Bradburd G.S. (2014) Isolation by environment. *Molecular Ecology*, **23**, 5649-5662.
Wang I.J., Glor R.E., & Losos J.B. (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters*, **16**, 175–182.
Waring G. & Price P. (1989) Parasitoid pressure and the radiation of a gallforming group (Cecidomyiidae: *Asphondylia* spp.) on creosote bush (*Larrea tridentata*). *Oecologia*, 293–299.
Waring G. & Price P. (1990) Plant water stress and gall formation (Cecidomyiidae: *Asphondylia* spp.) on creosote bush. *Ecological Entomology*, 87–95.
Warren D.L. & Seifert S. (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Society of America*, **21**, 335–342.
Weber J.N., Bradburd G.S., Stuart Y.E., Stutz W.E., & Bolnick D.I. (2016) Partitioning the effects of isolation by distance, environment, and physical barriers on genomic divergence between parapatric threespine stickleback. *Evolution*, 1–15.
Weir B.S. & Cockerham C.C. (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
Weller H.I. & Westneat M.W. (2019) Quantitative color profiling of digital images with earth mover's distance using the R package colordistance. *PeerJ*, **2019**, 1–31.
Werner F. & Olsen A. (1973) *Consumption of* Larrea *by Chewing Insects.* US International Biological Program, Desert Biome, Logan, UT.
Williams A.G. & Whitham T.G. (1986) Premature leaf abscission: an induced plant defense against gall aphids. *Ecology*, **67**, 1619–1627.
Wilson J.S. & Pitts J.P. (2010) Illuminating the lack of consensus among descriptions of earth

history data in the North American deserts: a resource for biologists. *Progress in Physical Geography*, **34**, 419–441.
Wiuf C., Zhao K., Innan H., & Nordborg M. (2004) The probability and chromosomal extent of trans-specific polymorphism. *Genetics*, **168**, 2363–2372.
Wright S. (1943) Isolation by distance. *Genetics*, **28**, 114–138.
Yang T.W. (1967) Ecotypic variation in *Larrea divaricata*. *American Journal of Botany*, **54**, 1041.
Yang T. (1970) Major chromosome races of *Larrea divaricata* in North America. *Journal of the Arizona Academy of Science*, **6**, 41–45.
Yukawa J. & Rohfritsch O. (2005) Biology and ecology of gall-inducing Cecidomyiidae (Diptera: Cecidomyiidae). *Biology, Ecology, and Evolution of Gall-Inducing Arthropods.* (ed. by A. Raman, C.W. Schaefer, and T.M. Withers), pp. 273–304. Science Publishers, Enfield, NH.
Yukawa J., Uechi N., Horikiri M., & Tuda M. (2003) Description of the soybean pod gall midge, *Asphondylia yushimai* sp n. (Diptera: Cecidomyiidae), a major pest of soybean and findings of host alternation. *Bulletin of Entomological Research*, **93**, 73–86.
Zuravnsky K.N. (2014) *Understanding the roles of polyploidy and the environment on nordihydroguaiaretic variation in* Larrea tridentata (Unpublished master's thesis). Salisbury University, Salisbury, MD.