Using Species Distribution Models to Predict Genomic Isolation: A Case Study with a High-altitude, Stream-specialist Damselfly (Odonata, Calopterygidae: *Hetaerina vulnerata* Hagen in Selys, 1853) in Arizona, New Mexico, and Utah

by

Austin R. Biddy, B.S.

A Thesis

In

Biology

Submitted to the Graduate Faculty of Texas Tech University in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

Approved

Nancy E. McIntyre, Ph.D.
   Chair of the Committee

Joseph D. Manthey, Ph.D.

Jessica L. Ware, Ph.D.

Mark Sheridan, Ph. D.
   Dean of the Graduate School

August, 2022
ACKNOWLEDGMENTS

Foremost, I would like to acknowledge and thank my advisor Dr. Nancy McIntyre. She has been a wonderful and patient mentor as I navigated this program. She always took time to chat over coffee and discuss science, research, or just great recipes and restaurants to try out. I am unequivocally thankful for her dedication to this thesis. Also, my committee members Drs. Joseph Manthey and Jessica Ware provided indelible insight and contributions to my thesis. This thesis was my first exposure to genomics, and their guidance and advice was more than necessary for its completion. I would also like to thank Dr. Pierre Deviche for his assistance with field work and collecting *Hetaerina vulnerata* in Arizona. This project would not have been completed without funding awarded from the Texas Tech Graduate School, the Dragonfly Society of the Americas, the Society of Integrative and Comparative Biology, Sigma Xi, and the Entomological Society of America.

Throughout my time at Texas Tech, I met wonderful friends and colleagues who were constant moral support. I am beyond thankful for my fellow lab mates: Sean Sutor, Danielle Husband, Joe Girgente, Hannah Wojtysiak, and Dr. Lucas Heintzman. In particular, Danielle and I shared an office, many laughs, and much needed movie nights prior to the pandemic; she also assisted with field work for both summers I collected *H. vulnerata*. I also want to extend an appreciation to Drs. Michael Dini, Carrie Monjé, Stephanie Lockwood, and Emma Roberts; they were all wonderful teaching mentors.

Of course, more on the home front, my parents, Linda and Robert Biddy, were always a phone call away, and I am thankful that I could call and chat every week or so to decompress and just talk. Also, my late grandmother instilled a love and passion for education that she was never able to receive, and she was a source of inspiration when this whole experience felt too daunting. My partner Adam Endreson has been a constant companion and always there to listen to what seemed like gibberish when I would belabor him with details of why I could not figure out a stat or program. I am even more thankful for the amazing meals he would cook for us, which were always a highlight at the end of a long, stressful day (whereby, I experienced more of those than I would like to admit!). Our cat Binx (who unfortunately passed away in April 2022) was a wonderful
and cuddly pet, definitely providing catharsis when it was needed the most. Our pug Chanel has been a source of bountiful energy, cuddles, and constant snorting and snoring. All the aforementioned people (and more!) made immense impacts on me as I have meandered through my thesis.
# TABLE OF CONTENTS

**ACKNOWLEDGMENTS** .................................................................................................................. ii

**ABSTRACT** .................................................................................................................................... v

**LIST OF TABLES** .......................................................................................................................... vi

**LIST OF FIGURES** ......................................................................................................................... vii

I. Using Species Distribution Models to Predict Genomic Isolation: A Case Study with a High-altitude, Stream-specialist Damselfly (Odonata, Calopterygidae: *Hetaerina vulnerata* Hagen in Selys, 1853) in Arizona, New Mexico, and Utah ................................................................................................................................. 1

 Introduction ........................................................................................................................................ 1

 Methods ............................................................................................................................................ 6

 Model species ................................................................................................................................... 6

 Species Distribution Models ............................................................................................................. 7

 Sampling localities ............................................................................................................................ 10

 DNA extractions and sequencing ....................................................................................................... 11

 Data analyses .................................................................................................................................... 13

 Results ............................................................................................................................................. 16

 Species Distribution Models ............................................................................................................. 16

 Genomic analyses ............................................................................................................................. 16

 Discussion ........................................................................................................................................ 19

 **BIBLIOGRAPHY** ......................................................................................................................... 25

 **APPENDICES** ............................................................................................................................... 46

 A. Supplementary table 1.1 ............................................................................................................. 46
Aquatic animals in high-elevation streams of the arid southwestern U.S. are potentially isolated by distance (IBD) or environment (IBE). The Canyon Rubyspot damselfly (*Hetaerina vulnerata*) is an insect that inhabits mountain streams in the American southwest. Spatial separation of streams and limited dispersal capacity of *H. vulnerata* may cause population isolation and genomic structuring, and projected climate change may exacerbate isolation by restricting distribution and movement. MaxEnt was used to construct a species distribution model (SDM) based on environmental variables and occurrences of *H. vulnerata*; this model indicated seven potential population clusters isolated by intervening unsuitable habitat. Projecting to future climate conditions, an overall gain of suitable habitat (0.8% increase) was identified, but current suitable habitat in southern Utah and northern New Mexico was lost. I collected 124 *H. vulnerata* from eight localities at six of the seven clusters; DNA was extracted from 78 individuals for ddRADseq genotyping. A suite of population and landscape genomic analyses were used: Admixture, IBD, F$^\text{ST}$, and a genotype-environment association (GEA). Admixture analyses determined six subpopulations, partially corroborating the SDM. IBD was not significant; IBE was only significant when considering two ecological variables. F$^\text{ST}$ values were low (< 0.07). GEA determined that 20 SNPs were locally adapted to tree canopy coverage. These results indicate that *H. vulnerata* populations are likely recently separated but undergoing genomic isolation and local adaptation. Integrating SDMs with landscape genomics combines techniques to indicate populations that are separated by distance and unsuitable habitat, providing explanations for patterns in genomic structuring.
LIST OF TABLES

1.1 Putative populations visited by watershed and associated mountain range. Elevation is in meters asl, and coordinates are N latitude / W longitude in decimal degrees; n corresponds to the number of individuals collected per watershed (total n = 124) ....................................................34

1.2 Mean depth coverage per putative population, as reported by the ustacks program. The number of individuals sequenced per locality is indicated by n ..........................................................35

1.3 Mean F_{IS} values per locality visited. 1 indicates inbreeding (homozygosity) and -1 a lack of inbreeding (heterozygosity) ..................................................................................................................36

1.4 Mean F_{ST} values per locality sampled, ranging from 0 (panmictic) to 1 (completely different) .................................................................................................................................37

1.5 Mantel’s test outputs for IBD and IBE. “All” corresponds to the test using every environmental variable. Each of these variables was then parsed out to determine which environmental constraint is likely impeding gene flow or movement across the landscape for H. vulnerata based on genetic distances. See Supplementary Table 1.1 for the description of these variables. The asterisks denote significant (p < 0.05, **) and marginally significant values (0.10 > p > 0.05, *) ..................................................................................................................38
LIST OF FIGURES

1.1 A) *H. vulnerata* photo (photo: N.E.M.); B) range map for *H. vulnerata* from the southwestern U.S. to Honduras based on iNaturalist observations (dark points).................................39
1.2 SDM output with seven putative populations outlined in black circles. Cool colors (e.g., blue) indicates low habitat suitability; warm colors (e.g., red) indicates high habitat suitability. White dots are occurrence points ...........................................40
1.3 Sampling locations in the National Forests (NF) of the southwestern U.S. during the summers of 2020 and 2021. Inset maps are provided for Zion National Park and Coronado NF (black bounding box) to display locations that were proximal to each other and difficult to discern from the overall map extent ....................................................................................41
1.4 Scree plot of environmental predictor variables; the first four principal components contributed to 76.4% of variance explained with these variables .................................................................42
1.5 Ancestry matrix (K = 6) of *H. vulnerata* subpopulations sampled across their extent. Each color corresponds to a population and its shared loci (ancestry proportions).................................43
1.6 Isolation by distance plot: x-axis is the geographic distance in decimal degrees and the y-axis is population-level F<sub>ST</sub>. The red line is a smoothed local mean, indicating that the relationship is not linear (Mantel’s test: p-value > 0.05). Warm colors represent high point density, cool colors represent low point density ..............................................................................44
1.7 Isolation by environment plots: x-axis is the environmental distance using Canberra’s distance and the y-axis is population-level F<sub>ST</sub>. The red line is a smoothed local mean, indicating that the relationship is not linear. Significant and marginally significant sets of variables and/or individual variables were plotted. Warm colors represent high point density, cool colors represent low point density.................................................................45
CHAPTER I

USING SPECIES DISTRIBUTION MODELS TO PREDICT GENOMIC ISOLATION: A CASE STUDY WITH A HIGH-ALTITUDE, STREAM-SPECIALIST DAMSELFLY (ODONATA, CALOPTERYGIDAE: HETAERINA VULNERATA HAGEN IN SELYS, 1853) IN ARIZONA, NEW MEXICO, AND UTAH

INTRODUCTION

Spatial fragmentation of populations has genomic consequences and can be generated by multiple factors. Isolation by distance (IBD) indicates that genomic similarity is a function of geographic distance between individuals’ localities. IBD is considered a null model for genomic structuring under the Wright (1943) conjecture and as such can be used to test if geographic separation is affecting genomic structuring, or if there are other extrinsic factors driving genomic patterns (e.g., isolation by environment [IBE], isolation by resistance [IBR], isolation by barrier [IBB] etc.). Because biotic presence is predicated at least in part by abiotic conditions, climate change is projected to affect the abundance and distribution of certain organisms by further isolating some types of habitats, because environmental conditions will be altered. Certain ecosystems may be especially sensitive and vulnerable to such changes, notably aquatic systems at high altitudes that are already isolated by drier, lower-elevation areas; these ecosystems and the populations they support are likely to experience unprecedented warming, evaporation, and loss of niche space in the future (Peeters et al., 2002; Beever et al., 2010; Angert et al., 2011; Ohmura, 2012; Diaz et al., 2014; Urbani et al., 2017; Birrell et al., 2020; Miller et al. 2021). The southwestern United States in particular has a high vulnerability to climatic change that is compounded by the inherent aridity of the region (Chylek et al., 2014). The southwestern U.S. has experienced a temperature increase over
the last several decades (estimated at 0.45°C since the mid-20th century; Van Devender and Brusca, 2015). Future climatic models suggest that increased temperatures, moderately decreased precipitation, and reduced streamflow are likely in the southwestern U.S. by the end of this century (Miller et al., 2021). These changes are anticipated to have biotic consequences (Van Devender and Brusca, 2015).

Since climate change may potentially exacerbate spatial isolation by affecting habitat availability and suitability, it is imperative to quantify current habitat suitability. Species distribution models (SDMs) identify the abiotic and biotic conditions (e.g., temperature, elevation, precipitation, tree canopy coverage, etc.) at species’ occurrence locations to identify areas where a species may occur but has not yet been recorded and areas that are currently unsuitable but that may become so via climate change or vice versa (Beck, 2013; Zhu et al., 2013; Urbani et al., 2017). In addition to modelling the current distribution of a species, SDMs can be projected to future conditions with a variety of different future scenarios or potential future outcomes to explore how habitat suitability may change. SDMs can determine where these vulnerable areas occur (Betts et al., 2014; Bosso et al., 2018).

For example, vulnerability of aquatic species in xeric areas such as the southwestern U.S. raises concerns about range truncations and the likelihood of extirpation in future climates, as suggested by projected SDMs (Urbani et al., 2017). Aquatic insects are poikilothermic and inherently susceptible to fluctuations in temperature and water availability (Jackson et al., 2020), and lotic species with somewhat limited dispersal abilities within mountainous streams and adjacent habitats in otherwise
xeric areas should especially exhibit population genetic structuring (Johnson, 1973; Grewe et al., 2012; Múrria et al., 2019). One group that may be especially sensitive to current conditions and to how climate change may affect the spatial distribution of discrete habitats is the order of semi-aquatic insects Odonata (dragonflies and damselflies) (Hassall, 2015; Bybee et al., 2016). Odonates are hemimetabolous, with the egg and naiad stages strictly aquatic and the adult stage aerial/terrestrial. Because of this life history, odonates are sensitive to habitat changes and drought; indeed, odonates are considered an important group of wetland indicators (Oertli, 2008). Most odonate species exhibit some habitat specificity (i.e., are either lentic or lotic), with some species-specific distributions associated with elevation (Ball-Damerow et al., 2014). This habitat specificity compounded with habitat isolation and vulnerability to climate change suggest that odonates will be susceptible to population isolation resulting from gene flow between fragmented populations being severed (Bush et al., 2014; Bellis et al., 2021). Because of their relatively low vagility, damselflies in isolated habitat patches that are vulnerable to climate change—such as high-elevation streams in the arid southwestern U.S.—are candidates with which to study genetic isolation and structuring (Lorenzo-Carballe et al., 2015).

_Hetaerina vulnerata_ Hagen in Selys, 1853 (Canyon Rubyspot damselfly; Fig. 1.1a) inhabits shaded streams in high-altitude areas (550-2000 m), ranging from Honduras up to the southwestern U.S., where it is found in high-elevation portions of Arizona, New Mexico, and Utah (Johnson, 1973; Alcock, 1982; Stevens and Bailowitz, 2009; Paulson, 2010) (Fig. 1.1b). It is considered a relatively slow flyer (taking longer to
thermoregulate and take flight) with limited dispersal abilities compared to *H. americana* (Rivas et al., 2016). Even though *H. vulnerata* occupies lotic waters, the streams of the southwestern U.S. experience episodic droughts and are isolated by intervening xeric areas, which may disrupt gene flow. Spatial separation of stream drainages combined with the limited dispersal capacity of *H. vulnerata* may induce population isolation and genetic structuring. Climate change should exacerbate population isolation in this species, potentially resulting in genomic structuring. Using SDMs, we should be able to identify current habitat for *H. vulnerata* that may indicate genomic patterns of differentiation or structuring related to geographic distance or availability of suitable environmental conditions.

Because SDMs map landscape suitability, they can also be used to describe potential gene flow between populations (Wang et al., 2008; Brown, 2014; Park et al., 2021; Biddy and McIntyre, in review). Dragonflies and damselflies have been the foci of numerous SDM studies (see review by Collins and McIntyre, 2015). Several studies employing SDMs have indicated projected climate change effects on odonates, including geographic isolation, hindered dispersal by loss of habitat, and reduced habitat suitability (Bush et al., 2014; Amundrud et al., 2017; Collins and McIntyre, 2017; Bellis et al., 2021). Because dispersal along streams is contingent upon water availability, the potential desiccation of streams could possibly impede gene flow in *H. vulnerata*, and SDMs can indicate such areas of sensitivity (Wang et al., 2008).

Although insects are highly diverse and ecologically important in aquatic and terrestrial ecosystems, there have been few population and landscape genomic studies
conducted on them, and even fewer on non-model insects (Wachi et al., 2018) such as Odonata (but see Ioannidis et al., 2017; Dudaneic et al., 2018). Genetic and genomic research on Odonata has been conducted to assess their phylogeny with mitochondrial DNA (mtDNA), microsatellites, and/or nuclear genes (e.g., Abbott et al., 2018; Torres-Cambas et al., 2019; Vega-Sánchez et al., 2019), or to assess population genetic structure and differentiation, typically with mtDNA (e.g., Lorenzo-Carballa et al., 2015; Low et al., 2017; Xue et al., 2017), with both mtDNA and nuclear DNA (e.g., Hayashi et al., 2005), or with whole genomic DNA (e.g., Watts et al., 2007; Dudaniec et al., 2018). However, few genomic studies on odonates have explicitly considered the effects of landscape patterns on genomic processes (Dudaneic et al., 2018), and landscape genomic studies have yet to be conducted on *H. vulnerata* across its U.S. range.

I used locality records of *H. vulnerata* to build a species distribution model under current and projected future climates to quantify changes in potential niche space, identify possible climate refugia, and explain any current patterns of genetic structure. Because *H. vulnerata* are high-altitude, stenotopic damselflies with populations in isolated, montane areas surrounded by lowland desert in the southwestern U.S., I hypothesized that these populations would differ in genetic structure shaped by isolation by distance, isolation by environment, or genetic clustering by mountain range. Due to a lack of connectivity from the paucity of water in this region, local adaptation was hypothesized to occur within and between populations. My thesis research will elucidate a gap in our knowledge concerning population and genomic isolation of *H. vulnerata* across its U.S. range and how this species might be affected by future climate change.
METHODS

Model species

*Hetaerina* (rubyspots) is the most speciose genus (39 spp.) of Calopterygidae in the New World (Garrison et al., 2010; World Odonata List: https://www2.pugetsound.edu/academics/academic-resources/slater-museum/biodiversity-resources/dragonflies/world-odonata-list2/) and is most diverse in South and Central America. There are only four species of this charismatic genus in the United States: *H. calverti* sp. nov. (Cryptic Rubyspot), *H. titia* Drury, 1773 (Smoky Rubyspot), *H. vulnerata* (Canyon Rubyspot), and *H. americana* Fabricius, 1798 (American Rubyspot). *H. calverti* was only recently described, so its full range is enigmatic, as it was previously considered a member of the *H. americana* species complex (Vega-Sánchez et al., 2020). Of the remaining three species, *H. americana* has the most widespread distribution in the U.S. whereas *H. vulnerata* has the most restricted range because of its habitat specificity (Garrison, 1990; Stevens and Bailowitz, 2009).

*Hetaerina* typically inhabit small streams and rivers, and certain species, such as *H. vulnerata*, occupy shaded streams and rivers. These species use shaded riparian areas for thermoregulation, as nighttime refugia, and for post-copulatory resting (Garcia-Garcia et al., 2017; Cordoba-Aguilar and Rocha-Ortega, 2019). *Hetaerina* males are often found in the same stretch of stream (Alcock, 1982), while females' position in the stream is not as predictable (Bick and Sulzbach, 1966). Fine-scale microhabitat preference makes *H. vulnerata* an emblematic species for studying distribution modeling and population isolation effects at the genomic scale.
Species distribution models

I constructed SDMs for *Hetaerina vulnerata* in the U.S. portion of its range, representing the northernmost (and presumably most recent) segment colonized; given that poleward movement is a climate-adaptation response (Parmesan et al., 1999). Models constructed to examine only this portion of the species’ range should be robust because of the widespread data available for North America (occurrence points, environmental data) relative to the rest of the species’ distribution. The SDMs were constructed using MaxEnt software. MaxEnt uses maximum entropy theory for species distribution estimations and projecting future occurrences. Models were developed with presence data from spatial coordinates and a suite of environmental data layers (Phillips et al., 2006). For *H. vulnerata* occurrence data, only verified species identification data from OdonataCentral (https://www.odonatacentral.org), the Global Biodiversity Information Facility (https://www.gbif.org), iNaturalist (https://www.inaturalist.org), and the Arizona Dragonflies (http://azdragonfly.org/) databases were used. Points from iNaturalist were filtered out if they were not verified by the community or if the approximate location range was imprecise (greater than 200 m). All 19 bioclimatic variables in the WorldClim-Global Climate version 2 data repository (Fick and Hijmans, 2017) were retrieved. Elevation and flow direction data (which is a slope raster that takes movement of stream drainages into account) were retrieved from USGS HydroSHEDS (https://hydrosheds.cr.usgs.gov), and the elevation data were used to create a slope layer. Because Canyon Rubyspots occupy shaded streams, tree canopy coverage (TCC) data
were retrieved from the U.S. Forest Service (https://data.fs.usda.gov/geodata/rastergateway/treecanopycover/). Grain resolution was at 30 arc-seconds for all layers. Collectively, using both abiotic and biotic data in the SDMs provides the best likelihood of accurately modeling the species’ current occurrence.

To project the SDM to future climate conditions (year 2080), data from WorldClim under varying general circulation models (GCMs) and greenhouse gas Representative Concentration Pathways (RCPs) were downloaded under the Climate Model Intercomparison Project Phase 5 (CMIP5) at a resolution of 30 arc-seconds. For GCMs, I selected the IPSL-CM5A-MR and MIROC-ESM models, representing projected conditions of aridity and increased temperature and average precipitation rates and extreme temperature increases, respectively. The RCPs selected were from the high end of the range for both GCMs, RCPs 6.0 and RCPs 8.5. This protocol brackets probable climate futures for the southwestern U.S. (Miller et al., 2021). Similar to the current data, I quantified future distribution of *H. vulnerata* with the 19 bioclimatic variables, elevation, slope, and flow direction. Tree canopy coverage was excluded, as there are no estimated future data for TCC for the southwestern U.S. Variable names and abbreviations can be found in Supplementary Table 1.1.

I used a Principal Component Analysis (PCA) ordination method to determine environmental variables that are contributing to overall model variance in R 1.4.1103 (R Core Team, 2019) for current projections, and these variables were used for future scenarios to maintain variable consistency between models. For the SDM under current
climate conditions, I used a PCA to determine model variance of the 23 variables. The first four PC axes explained 76.4% of model variance (Fig. 1.4); cumulative variance was negligible for the remainder of the axes. To select variables, I examined loadings (scaled eigenvalues). Assessing the loadings of PC1, the variable loadings that were higher than those in PC2 were selected, and this process was continued for PC2 through PC4. Thus, there was overlap in variable selection. However, lower loading scores of each subsequent axis were omitted from selection, due to low variance of the predictor. The variables from each PC axis were subset into respective folders for kuenm (Cobos et al., 2019) to query and build candidate models. After testing and evaluating calibration models (n = 124), the model with PC1 variables (Bio1, Bio5, Bio10, Bio12, Bio17, DEM, TCC [Supplementary Table 1.1]) fit the significance criteria best. This model had a regularization multiplier of 1 and with a linear and quadratic feature applied was closest to the significance criteria (AICc = 5275, OR = 6.5%). The final model was generated, based on an average of ten replicate SDMs.

In the package kuenm, I built and evaluated candidate models of SDMs, selecting an SDM that was the most parsimonious (low AICc) and had the lowest omission rate (OR < 0.05). Models were tested and trained with a subset of coordinates. When the final model was created, 10 SDM outputs were plotted; an average of these replicates was used for determining sampling localities and comparing between climate regimes (i.e., the final model was based on an average of ten replicate SDMs). The current SDM was used to identify geographically discrete locations with putatively isolated populations (Fig. 1.2), which were then sampled for genomic analyses (Fig. 1.3).
Sampling localities

Based on the SDM output, seven *H. vulnerata* population clusters across the study extent were estimated to be present, indicated by clustering of occurrences and presence of unsuitable habitat separating populations. From each of the seven putative clusters, multiple sampling locations were visited per cluster (Table 1.1). No damselflies were located in either year at one of the population clusters (the Santa Rita Mountains and Huachuca Mountains) as most streams were completely dry; from the remaining five, I collected a minimum of 10 specimens (except for Dixie National Forest in Utah, where only \( n = 9 \) were able to be located) as an adequate number to assess genomic structure via ddRAD-Seq (Yadav et al., 2019). Based on known occurrences from citizen-science repositories within the putatively isolated populations identified by the SDM, sampling locations were chosen along shaded streams, a known microclimate preference. I focused on adults rather than naiads because of a lack of reliable taxonomic keys for identification, particularly for very early instars. Adults were collected with aerial insect nets and immediately preserved in 100% ethanol. Collecting was performed during times of high damselfly activity, late morning to midday during the post-monsoon season from July-September (Stevens and Bailowitz, 2009). I collected in the summers of 2020 (\( n = 101 \)) and 2021 (\( n = 23 \)) (Fig. 1.3). Based on the SDM, Lincoln National Forest (New Mexico), the Jemez Springs area in Santa Fe National Forest (New Mexico), and Pine Valley in Dixie National Forest (Utah) were indicated as areas of high habitat suitability, yet presence of *H. vulnerata* in those areas had not been documented. I sampled at these localities to determine species presence but did not detect any *H. vulnerata* (possibly due
to limited sampling time, recent changes in habitat suitability due to forest fires, or over-estimation predicted by the model). A total of 124 individuals were collected across summer 2020 and 2021. From each population cluster/locality sampled (n = 8), 10 individuals were selected for DNA extraction. Since I did not sample 10 from Dixie National Forest, only 9 were used for extraction and sequencing. After sequences were received and analyzed, a likely *H. americana* was detected from the Coconino National Forest population and was removed for subsequent analyses. Thus, only 78 individuals were successfully sequenced (Gila National Forest = 10, Coronado National Forest [Chiricahua Mts and Pinaleño Mts.] = 20, Santa Fe National Forest = 10, Tonto National Forest = 10, Coconino National Forest = 9, Utah [Dixie National Forest and Zion National Park = 19).

**DNA extraction and sequencing**

Specimens were transported back to Texas Tech University for DNA extraction. Once completely air dried, whole genomic DNA was extracted from the thoracic muscle (Watts et al., 2007; Ioannidis et al., 2017) of 78 *H. vulnerata* using the Qiagen (Germantown, MD) DNeasy Blood and Tissues kit, abiding by the protocol of the manufacturer for insect tissue, in which the thoracic tissue was pulverized by mortar and pestle in liquid nitrogen, then 180 µl of Buffer ATL was added. After the addition of 20 µl of proteinase K, the sample was vortexed and subsequently incubated at 56°C until lysing was completed. This process took three hours, and the samples were periodically vortexed. Then 200 µl of Buffer AL and 200 µl of 100% ethanol was mixed and added to the sample, vortexing the homogenate before and after addition of this mixture. This
entire mixture and precipitate were placed into a 2 ml collection tube in a DNeasy Mini spin column and centrifuged at 8,000 rpm for 1 minute. The flow-through (the liquid that distills beneath the collection tube) and collection tube were removed. The DNeasy Mini spin column was inserted into another, clean 2 ml collection tube with 500 µl of Buffer AW1, centrifuged for 1 minute at 8,000 rpm, and once again the flow-through and collection tube were discarded. The spin column was inserted into another 2 ml collection tube, adding 500 µl of Buffer AW2, and centrifuged for 3 minutes at 14,000 rpm, which evaporated any residual ethanol that could corrupt later reactions. The flow-through and collection tube were disposed, and the dry spin column was carefully removed so that it did not touch or interact with any ethanol. The spin column was placed into a 2 ml microcentrifuge tube, then 200 µl of Buffer AE was directly pipetted onto the DNeasy membrane. After incubating at room temperature for 15-20 minutes to increase yield, the sample was centrifuged for an additional minute at 8,000 rpm, eluting the sample. DNA concentration was quantified via a Qubit Fluorometer. Subsequently, the whole-genome extracts were sent for sequencing and library preparation via Illumina HiSeq at Admera Health, LLC (South Plainfield, NJ).

The method for sequencing (ddRAD-seq; Peterson et al., 2012) used two restrictive enzymes (NlaIII-MluCl), which randomly excised portions of the genome; genomic diversity and single nucleotide polymorphisms (SNPs) were estimated from these excised regions of the genome (Davey et al., 2011; Andrews et al., 2016). This method does not require a reference genome (unavailable for H. vulnerata), and the enzyme combination that was used yields ~100K size-selected fragments (Peterson et al.,
2012). Relatively low-cost and expeditious, ddRAD-seq only requires approximately 100 ng of DNA for analyses on wild specimens. In addition, several studies have successfully utilized ddRAD-seq on non-model insects and other arthropods (Fritz et al., 2016; Lal et al., 2016; Kozlov et al., 2017; Lado et al., 2019; Yadav et al., 2019) on various topics (e.g., population genetics, phylogenetics, inbreeding, etc.).

**Data analyses**

Raw sequences were filtered and aligned using the software Stacks with default settings and a phred quality scoring of 10 (Catchen et al., 2013). Default parameters were used in Stacks, aside from some slight modifications of loci detected in ustacks (Table 1.2) and cstacks. The parameter for maximum distance permitted between nucleotides between each stack, which is implemented in ustacks, was adjusted to 4 (default = 2), and the same value was retained in cstacks for number of mismatched loci allowed between sample loci while the catalog was being constructed, as recommended by Catchen et al. (2013). In the populations program, I selected 1 for the minimum number of populations a locus must be present in to process a locus. Minimum percentage of individuals in a population required to process a locus for that population was 90% for a robust representation of each population, and linked SNP inclusion was minimized by only selecting the first SNP per locus. Based on a quality scoring of phred = 10 for process_radtags filtering of fastq files, 99.4% of reads were retained. Following Stacks processing, I retained 9,292 loci and 6,054 SNPs. Stacks output files (e.g., VCF, Genepop, Structure, etc.) were used for all subsequent analyses.

Since individuals collected from a few of the putative populations were from the same stream catchment, I used the *related* package (Pew et al., 2015) and FIS via *hierfstat*
(Goudet, 2021) in R to quantify within-population variation. The related package determines combinations of coancestry and whether individuals are parent-offspring, siblings, or cousins via relatedness estimates by Li et al. (1993) and Wang (2002). Pairs that were determined as parent-offspring (coancestry > 0.5) were eliminated from other analyses. F_{IS} ranges from -1 (complete heterozygosity) to 1 (complete homozygosity), thus providing an estimate of inbreeding within populations.

F_{ST} is a metric to indicate degree of genetic differentiation between populations that is scaled from 0 (panmixia) to 1 (complete differentiation). I used the hierfstat package using the default setting’s Cavalli-Sforza and Edwards Chord distance, which assumes genetic drift as the evolutionary agent acting on populations. IBD and IBE were determined using the Mantel test in vegan (Oksanen et al., 2020) in R. Geographic distances were in decimal degrees, and for IBE, a Canberra’s distance was applied to the environmental data using all 23 environmental variables (i.e., bioclimatic data, tree canopy coverage, elevation, etc.). A Markov Chain Monte Carlo with Cross Validation (MCMCCV) was used to simulate whether geographic distance, environmental distance, or both distance matrices were contributing to genetic variance, determined with a likelihood value using the Sunder package in R (Bradburd et al., 2013). Simulations of distances with the lowest likelihood values were considered as contributing toward genetic distance. I used the package LEA (Frichot and François, 2015) in R for a numerical optimization algorithm of mean-square estimates of ancestry proportions (which assumes that the populations are in Hardy-Weinberg Equilibrium) with a sparse Non-Negative Matrix Factorization (sNMF) method to determine the ancestry coefficient.
and to plot clustering of putative populations. Furthermore, a PCA was used to model population clustering by visualizing the relationships of genotypes in ordination space. Thus, I was able to corroborate the number of \textit{a priori}, putative populations inferred from the SDM. I assigned the K values as 1-8 accounting for putative populations inferred from the SDM (n = 7) and potential within-population structuring; to estimate the number of subpopulations, I configured the sNMF function with 50 repetitions per K, 1,000,000 iterations per K, and an alpha of 10 (regularization parameter). The model with the K that had the lowest cross-entropy value across repetitions was suggested as the “true” K.

I used a genotype-environment association as a statistical analysis to determine local adaptation across these high-elevation systems (Lotterhos and Whitlock, 2015). This was estimated with a univariate latent factor mixed model (LFMM) with the \textit{lfmm} package (Frichot et al., 2013) and multivariate redundancy analysis (RDA) with the \textit{vegan} package (Oksanen et al., 2020). I used four variables for the GEA (both LFMM and RDA), based on ecological pertinence for \textit{H. vulnerata} (Bio10, Bio13, Elevation, TCC). Genotypes were imputed for the LFMM analysis, and a matrix of environmental variables was used to test for genotype-environment associations. Using a multivariate ordination approach to quantifying local adaptation takes the multivariate environment into account, rather than just using a univariate method alone (i.e., LFMM) (Forester et al., 2018).
RESULTS

Species distribution models

Considering future climate scenarios and emission rates, only one model was significant when evaluated: GCM MIROC-ESM and RCP 6.0 with an AICc = 5812 and OR = 4.3%. The variables used for model construction from PC1 under current climate projections were maintained for the future model, except for tree canopy coverage, which has not been projected for this portion of the country. The regularization multiplier for the model was 1, and a quadratic feature was applied to the model. Although this represents a marked increase in suitable habitat under future projections for *H. vulnerata*, the available habitat was more clustered and less dispersed than the current distribution. The loss of current suitable habitat for *H. vulnerata* was principally in the northernmost portion (Utah and New Mexico) of the species’ range in the future model. Suitable habitat was gained near the Wasatch range of Utah and the San Juan Mountains of Colorado (Fig. 1.2).

Genomic analyses

As determined via the related analyses, individuals collected within populations were not parent-offspring or siblings (coancestry < 0.5). $F_{IS}$ values showed no trend in values based on geographic proximity or distance between populations.

Therefore, populations with negative inbreeding coefficients had higher heterozygosity values than those with positive inbreeding coefficients (Table 1.3) Using the Admixture analysis with a K-value from 1-8, the different populations began to parse out from one another at K = 6 (Fig. 1.5). The lowest cross-entropy value (cross-entropy =
0.2232) indicated that the “true” number of subpopulations was 6 (K = 6). Loci were shared between all populations, particularly in the Dixie, Pinaleño, and Zion localities. The FST values were relatively low (< 0.07), but the Chiricahua Mts. cluster had markedly consistent genetic differentiation between putative clusters, and this is irrespective of proximity between populations (Table 1.4). Furthermore, the genetic differentiation of the putative population from Dixie National Forest and Zion National Park (both in Utah) demonstrates similarity by proximity.

IBD quantifies whether genetic distance is a proxy of geographic distance rather than other landscape factors, whereas IBE takes landscape and environmental factors into account for genetic distance. Using a Mantel’s test, the p-value was not significant for IBD. Since this is not a significant value, the Wright model of IBD is rejected, suggesting other models of differentiation (such as IBE) across the landscape for H. vulnerata (Figs. 1.6 and 1.7). However, the Mantel’s test for IBE was not significant. Analyzing each of the environmental variables, all were not significantly associated with genetic distance (Table 1.5). Cumulatively and individually, environmental variables did not impose a constraint for H. vulnerata across the landscape. Thus, the temporal ephemerality of streams in the southwestern U.S. may affect isolation, rather than spatial separation. The MCMCCV simulation uses a likelihood value to examine whether geographic distance, environmental distance, or the combination of the two contributes to genetic isolation. With the lowest likelihood value of -40123.45, environmental distance was influencing genetic variation for H. vulnerata across the landscape, based on this simulation.
Using the univariate LFMM analysis with K = 6, there were 269 SNPs under selection to environmental variables (candidate SNPs) after using a false discovery rate (> 0.0001) to test for multiple testing correction, with 166 SNPs unique to this analysis. For the RDA, cumulative variance was found in the first four axes, when 23 environmental variables were used. Since 100% of model variance was explained in the first 4 RDA axes (regardless of variable combination), I selected four variables that were ecologically pertinent (Bio10, Bio13, elevation, and tree canopy coverage). Regarding the multivariate RDA, the number of candidate SNPs was 435 (all associated with TCC), with 269 unique SNPs to this analysis. Considering the LFMM and RDA, 20 candidate SNPs were shared and inferred to be associated with tree canopy coverage.
DISCUSSION

My study is the first to document how *H. vulnerata* serves as a model taxon for effects of habitat isolation and climate change, and how these factors are impacting the populations at the genomic level. These effects were *a priori* inferred from the constructed SDM in current and future projections. The putative population clusters identified by the SDM were partially corroborated via genomic analyses (*K* = 6); populations separated by distance and xeric habitat displayed genomic differences. Microclimate and habitat specificity of *H. vulnerata* likely contribute to this isolation, and projected climate scenarios should exacerbate these conditions (Fig. 1.2). Populations in the northernmost terminus of their U.S. range are anticipated to lose suitable habitat, which may lead to extirpation (Travis, 2003; Betts et al., 2014; Urbani et al., 2017; Bosso et al., 2018); habitat suitability for *H. vulnerata* will likely be found at higher altitudes than presently. At a more northerly latitude, the Wasatch Range (Utah) and San Juan Mountains (Colorado) are likely to be suitable for *H. vulnerata* under future conditions. However, there is minimal suitable habitat between current populations and the projected suitable habitat. Thus, it is unlikely that *H. vulnerata* will disperse and colonize those areas. If successful establishment of a population were to occur, individuals would be further isolated by distance and environment from the remaining *H. vulnerata* populations. Elucidation of high-elevation, aquatic insect taxa reacting to broad-scale (both spatial and temporal) climate change is a challenge (see review by Birrell et al., 2020). Nevertheless, high-altitude specialists seldom latitudinally disperse from mountains (Leach et al., 2015; Siefert et al., 2015), supporting the notion that
colonization of higher elevations is more probable (Chen et al., 2011). These features point towards diminishing suitable habitat for *H. vulnerata* with constriction by elevation.

Isolation by environment is caused by the unavailability of suitable habitat between isolated populations, which curtails migration and gene flow between populations (Wang and Bradburd, 2014). Since *H. vulnerata*, like most lotic specialists, disperse along watersheds (Grewe et al., 2012), the ephemeral streams of the southwestern U.S. are ostensibly the environmental factor isolating these populations. However, the mean temperature of the driest quarter (March – June for this region) and precipitation of the coldest quarter (November – February for this region) were the only significant environmental variables for IBE; I speculate that aridity and cooler temperatures are the isolating environmental variables for *H. vulnerata*. Based on these data, *H. vulnerata* is experiencing isolation by environment (as indicated by the MCMCCV simulation), as has been documented with other high-elevation species (Manthey and Moyle, 2015; Medina et al., 2021). However, milder temperatures during the autumn and winter in the southwestern U.S. is associated with the presence of some *H. vulnerata* at low altitudes (vetted observations on iNaturalist), but it is not yet known if these individuals are dispersing to other subpopulations because of a possible lack of structural connectivity between watersheds.

One anomaly in these data is that individuals sampled in the Dixie, Zion, and Pinaleño localities shared ancestry coefficients, per the Admixture analysis (Fig. 1.5), yet *F_{ST}* values suggest that the Pinaleño population is more closely related to individuals collected from Tonto and Coconino National Forests than to Dixie National Forest and
Zion National Park (Table 1.4). Although the Dixie, Zion, and Pinaleño populations are separated by distance and should thus reflect a classic IBD model (yet do not), I speculate that the populations are either undergoing similar selective pressures (Renaut et al., 2013) or have a large enough effective population size that is experiencing genetic drift at a slower pace.

Overall, $F_{ST}$ values were low but showed a relative trend based on geographic distance: the subpopulation in the Chiricahua Mountains consistently exhibited differences from other populations, irrespective of distance. Since $F_{ST}$ values were low, it is conceivable that there is gene flow, these subpopulations were only recently isolated after colonization of high-elevation habitat (Manthey and Moyle, 2015; Hill and Unckless, 2021), or are panmictic. However, odonates have been shown to exhibit slow evolutionary rates (Kjer et al., 2006). Therefore, the Chiricahua population is likely to be the most isolated at the genome level, and future climatic isolation should exacerbate this differentiation. The presence of such structuring within putative subpopulations suggests that although nearby subpopulations share some ancestral loci (Fig. 1.5), a lack of gene flow due to extrinsic landscape factors has potentially resulted in clustering (Medina et al., 2021). $F_{IS}$ values showed both heterozygosity and homozygosity, which could be a function of differences in assortative mating or effective population sizes. Future projections indicate that populations are likely to become more isolated, which may result in an increase of homozygosity and would be potentially deleterious for population persistence.
Although some organisms are expected to shift their range rather than be extirpated with climate change (Lyons et al., 2019), this is species-specific, as each taxon responds differently to habitat alterations (Betts et al., 2014; Cacciapaglia and van Woesik, 2018). For example, Bosso et al. (2018) speculated that low-elevation beetles should encroach on higher-elevation areas with future warming. This encroachment compounded with climate change is anticipated to negatively affect aquatic organisms at higher elevations (Slavich et al., 2014; Bellis et al., 2021; Timoner et al., 2021), such as *H. vulnerata*. Thus, it is imperative to consider local adaptation and phenotypic plasticity (Garzón et al., 2019). Moreover, even if suitable higher-elevation areas are present, habitat selection at a fine scale (based on whether a localized site provides appropriate resources) will dictate whether a population is able to persist there (Kirkton and Schultz, 2001; Hillman et al., 2014; Polic et al., 2014), but SDMs typically cannot capture such microhabitat details (Collins and McIntyre, 2015).

Regardless of the ecosystem of interest, most studies only consider bioclimatic or abiotic variables, but Forester et al. (2018) identified local adaptation to both abiotic and biotic variables in wolves, using univariate and multivariate models. Indeed, local adaptation studies have been applied to several high-altitude insects, e.g., Diptera (Hill and Unckless, 2021), Hymenoptera (Jackson et al., 2020), and Orthoptera (Tregenza et al., 2021). The findings of my study suggest (based on the shared candidate SNPs detected by the GEA) that *H. vulnerata* is undergoing local adaption to a biotic variable: tree canopy coverage. Interestingly, females of this genus submerge themselves underwater, actively ovipositing in tree root mats (Johnson, 1961; Alcock, 1982;
Eberhard, 1986; Biddy et al., 2020); presence of riparian trees is thus integral in their life cycle. Although at 30 arc-second resolution riparian trees cannot be distinguished from other forest types, there is a genotype-environment association to this variable. Under anticipated climate changes in the southwestern U.S., wildfires are expected to increase, imperiling tree populations (Friggens and Finch, 2015). Results presented from this study, using biotic variables, should thus elucidate population persistence for *H. vulnerata*, especially since local adaption is associated with a variable that may not occur in a commensurate density under future projections (Capblancq et al., 2020). In addition, the SDM under future projections included current tree canopy coverage distributions (Fig. 1.6); therefore, the future model is likely to be overestimated.

A reference genome for *H. vulnerata* does not currently exist, so the data from ddRADseq provides a reduced representation of the genome (Peterson et al., 2012). Therefore, patterns observed in the data do not reflect the entire genomic and demographic history of *H. vulnerata*. Future studies should endeavor to use whole genome and transcriptome data to increase the resolution of the analyses presented here, as well as test for selection and identify neutral and adaptive loci. This is a pressing need because of imminent climate change posing one of the most serious threats to biodiversity. These climatic changes have caused a decline in insect species and populations (see review by Wagner, 2020). The rate of climate change is outpacing the development of skills, tools, and techniques. My study employs a set of skills and techniques that researchers can use to identify genomic isolation and corroborate SDMs with genomic analyses considering projected climatic regimes (see also Biddy and
McIntyre, in review). In conjunction, SDMs and genomics are a powerful combination to examine other habitat specialists, providing insights into population persistence based on available habitat under future conditions as well as on populations likely to be extirpated due to unsuitable habitat compounded by maladaptation. *H. vulnerata* has served as an adequate model for the combination of using SDMs and landscape genomics, and this study exemplifies their utility in integrative studies.
BIBLIOGRAPHY


https://doi.org/10.1673/031.013.10201
Table 1.1: Putative populations visited by watershed and associated mountain range; n corresponds to the number of individuals collected per watershed (total n = 124).

<table>
<thead>
<tr>
<th>Population</th>
<th>Watershed</th>
<th>Mountain range</th>
<th>Elevation (meters asl)</th>
<th>Coordinates (N, W)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gila</td>
<td>Gallinas Creek</td>
<td>Black Mts.</td>
<td>2097</td>
<td>-32.8977, 107.8244</td>
<td>16</td>
</tr>
<tr>
<td>Chiricahua</td>
<td>Cave Creek</td>
<td>Chiricahua Mts.</td>
<td>1792</td>
<td>-31.8722, 109.2342</td>
<td>16</td>
</tr>
<tr>
<td>Pinaleño</td>
<td>Noon Creek</td>
<td>Pinaleño Mts.</td>
<td>1838</td>
<td>-32.6511, 109.8127</td>
<td>13</td>
</tr>
<tr>
<td>Santa Fe</td>
<td>Pajarito Spring</td>
<td>Jemez Mts.</td>
<td>1700</td>
<td>-35.8038, 106.1969</td>
<td>14</td>
</tr>
<tr>
<td>Tonto / Coconino</td>
<td>Horton Creek</td>
<td>Mogollon Rim</td>
<td>1713</td>
<td>-34.3458, 111.09</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Christopher Creek</td>
<td>Mogollon Rim</td>
<td>1710</td>
<td>-34.3088, 111.0383</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Wet Beaver Creek</td>
<td>Mogollon Rim</td>
<td>1161</td>
<td>-34.6736, 111.7081</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Oak Creek</td>
<td>Mogollon Rim</td>
<td>1658</td>
<td>-34.9958, 111.7378</td>
<td>11</td>
</tr>
<tr>
<td>Utah</td>
<td>Leeds Canyon</td>
<td>Pine Valley Mts.</td>
<td>1414</td>
<td>-37.2733, 113.3886</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Ash Creek</td>
<td>Pine Valley Mts.</td>
<td>1024</td>
<td>-37.2569, 113.2861</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Virgin River</td>
<td>Colorado Plateau</td>
<td>1301</td>
<td>-37.25, 112.95</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 1.2: Mean depth coverage all loci with variable sites per putative population, as reported by the ustacks program. The number of individuals sequenced per locality is indicated by n.

<table>
<thead>
<tr>
<th>Locality</th>
<th>n</th>
<th>Mean depth coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gila</td>
<td>10</td>
<td>7.93 ×</td>
</tr>
<tr>
<td>Coronado</td>
<td>20</td>
<td>9.76 ×</td>
</tr>
<tr>
<td>Santa Fe</td>
<td>10</td>
<td>7.70 ×</td>
</tr>
<tr>
<td>Tonto and Coconino</td>
<td>19</td>
<td>7.66 ×</td>
</tr>
<tr>
<td>Utah</td>
<td>19</td>
<td>9.39 ×</td>
</tr>
</tbody>
</table>
Table 1.3: Mean F<sub>IS</sub> values per locality visited. Positive values indicate inbreeding (homozygosity), whereas negative values represent an excess of heterozygosity.

<table>
<thead>
<tr>
<th>Population</th>
<th>F&lt;sub&gt;IS&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gila</td>
<td>0.175</td>
</tr>
<tr>
<td>Chiricahua</td>
<td>0.0557</td>
</tr>
<tr>
<td>Santa Fe</td>
<td>0.0462</td>
</tr>
<tr>
<td>Tonto</td>
<td>0.162</td>
</tr>
<tr>
<td>Coconino</td>
<td>-0.0719</td>
</tr>
<tr>
<td>Dixie</td>
<td>-0.188</td>
</tr>
<tr>
<td>Pinaleño</td>
<td>-0.441</td>
</tr>
<tr>
<td>Zion</td>
<td>-0.551</td>
</tr>
</tbody>
</table>
Table 1.4: Mean $F_{ST}$ values per locality sampled, ranging from 0 (panmictic) to 1 (completely different).

<table>
<thead>
<tr>
<th></th>
<th>Gila</th>
<th>Chiricahua</th>
<th>Santa Fe</th>
<th>Tonto</th>
<th>Coconino</th>
<th>Dixie</th>
<th>Pinaleño</th>
<th>Zion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiricahua</td>
<td>0.0614</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Santa Fe</td>
<td>0.0452</td>
<td>0.0669</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tonto</td>
<td>0.0239</td>
<td>0.0623</td>
<td>0.0411</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coconino</td>
<td>0.0232</td>
<td>0.0520</td>
<td>0.0222</td>
<td>0.0159</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dixie</td>
<td>0.0278</td>
<td>0.0506</td>
<td>0.0180</td>
<td>0.0240</td>
<td>0.0188</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pinaleño</td>
<td>0.0128</td>
<td>0.0579</td>
<td>0.0125</td>
<td>0.0138</td>
<td>0.0114</td>
<td>0.0530</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Zion</td>
<td>0.0259</td>
<td>0.0627</td>
<td>0.0194</td>
<td>0.0206</td>
<td>0.0222</td>
<td>0.00532</td>
<td>0.0381</td>
<td></td>
</tr>
</tbody>
</table>
Table 1.5: Mantel’s test outputs for IBD and IBE, with genetic distance as the response variable across all predictors. “All” corresponds to the test using every environmental variable. Each of these variables was then parsed out to determine which environmental constraint is likely impeding gene flow or movement across the landscape for *H. vulnerata* based on genetic distances. See Supplementary Table 1.1 for the description of these variables. The asterisks denote significant (p < 0.05, **) and marginally significant values (0.10 > p > 0.05, *).

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>Mantel statistic r</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geographic distance</td>
<td>0.193</td>
<td>0.195</td>
</tr>
<tr>
<td>Environmental distance</td>
<td>0.2864</td>
<td>0.098*</td>
</tr>
<tr>
<td>Bio1</td>
<td>0.7762</td>
<td>0.47</td>
</tr>
<tr>
<td>Bio2</td>
<td>0.2746</td>
<td>0.195</td>
</tr>
<tr>
<td>Bio3</td>
<td>0.2783</td>
<td>0.224</td>
</tr>
<tr>
<td>Bio4</td>
<td>0.1017</td>
<td>0.424</td>
</tr>
<tr>
<td>Bio5</td>
<td>0.71</td>
<td>0.464</td>
</tr>
<tr>
<td>Bio6</td>
<td>0.1872</td>
<td>0.325</td>
</tr>
<tr>
<td>Bio7</td>
<td>0.09182</td>
<td>0.428</td>
</tr>
<tr>
<td>Bio8</td>
<td>-0.0779</td>
<td>0.682</td>
</tr>
<tr>
<td>Bio9</td>
<td>0.5582</td>
<td>0.039**</td>
</tr>
<tr>
<td>Bio10</td>
<td>0.07506</td>
<td>0.482</td>
</tr>
<tr>
<td>Bio11</td>
<td>0.4512</td>
<td>0.63</td>
</tr>
<tr>
<td>Bio12</td>
<td>0.2101</td>
<td>0.23</td>
</tr>
<tr>
<td>Bio13</td>
<td>0.09931</td>
<td>0.414</td>
</tr>
<tr>
<td>Bio14</td>
<td>0.3257</td>
<td>0.139</td>
</tr>
<tr>
<td>Bio15</td>
<td>0.0747</td>
<td>0.468</td>
</tr>
<tr>
<td>Bio16</td>
<td>0.2991</td>
<td>0.162</td>
</tr>
<tr>
<td>Bio17</td>
<td>0.08476</td>
<td>0.425</td>
</tr>
<tr>
<td>Bio18</td>
<td>0.0174</td>
<td>0.47</td>
</tr>
<tr>
<td>Bio19</td>
<td>0.5149</td>
<td>0.04**</td>
</tr>
<tr>
<td>Elevation</td>
<td>0.05083</td>
<td>0.52</td>
</tr>
<tr>
<td>Flow direction</td>
<td>0.3591</td>
<td>0.095*</td>
</tr>
<tr>
<td>Tree canopy coverage</td>
<td>0.001698</td>
<td>0.484</td>
</tr>
</tbody>
</table>
Figure 1.1: A) *H. vulnerata* photo (photo: N.E.M.); B) range map for *H. vulnerata* from the southwestern U.S. to Honduras based on iNaturalist observations (dark points)
Figure 1.2: SDM output with six putative populations outlined in black circles. Cool colors (e.g., blue) indicate low habitat suitability; warm colors (e.g., red) indicate high habitat suitability. White dots are occurrence points.
Figure 1.3: Sampling locations in the National Forests (NF) of the southwestern U.S. during the summers of 2020 and 2021. Inset maps are provided for Zion National Park and Coronado NF (black bounding box) to display locations that were proximal to each other and difficult to discern from the overall map extent.
Figure 1.4: Scree plot of environmental predictor variables; the first four principal components contributed to 76.4% of variance explained with these variables.
Figure 1.5: Ancestry matrix (K = 6) of *H. vulnerata* subpopulations sampled across their extent. Each color corresponds to a population and its shared loci (ancestry proportions).
**Figure 1.6:** Isolation by distance plot: x-axis is the geographic distance in decimal degrees and the y-axis is population-level $F_{ST}$. The red line is a smoothed local mean, indicating that the relationship is not linear (Mantel’s test: p-value > 0.05). Warm colors represent high point density, cool colors represent low point density.
Figure 1.7: Isolation by environment plots: x-axis is the environmental distance using Canberra’s distance and the y-axis is population-level $F_{ST}$. The red line is a smoothed local mean, indicating that the relationship is not linear. Significant and marginally significant sets of variables and/or individual variables were plotted. Warm colors represent high point density, cool colors represent low point density.
Appendix A

Supplementary Table 1.1: Names of variables used for constructing the SDMs under future and current climate conditions; however, tree canopy coverage was not used for future conditions.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bio1</td>
<td>Annual mean temp</td>
</tr>
<tr>
<td>Bio2</td>
<td>Mean diurnal range (mean of monthly (max temp - min temp))</td>
</tr>
<tr>
<td>Bio3</td>
<td>Isothermality (Bio 2/Bio 7) (x 100)</td>
</tr>
<tr>
<td>Bio4</td>
<td>Temperature seasonality (standard deviation x 100)</td>
</tr>
<tr>
<td>Bio5</td>
<td>Max temperature of warmest month</td>
</tr>
<tr>
<td>Bio6</td>
<td>Min temperature of coolest month</td>
</tr>
<tr>
<td>Bio7</td>
<td>Temperature annual range (Bio 5 – Bio 6)</td>
</tr>
<tr>
<td>Bio8</td>
<td>Mean temperature of wettest quarter</td>
</tr>
<tr>
<td>Bio9</td>
<td>Mean temperature of driest quarter</td>
</tr>
<tr>
<td>Bio10</td>
<td>Mean temperature of warmest quarter</td>
</tr>
<tr>
<td>Bio11</td>
<td>Mean temperature of coldest quarter</td>
</tr>
<tr>
<td>Bio12</td>
<td>Annual precipitation</td>
</tr>
<tr>
<td>Bio13</td>
<td>Precipitation of wettest month</td>
</tr>
<tr>
<td>Bio14</td>
<td>Precipitation of driest month</td>
</tr>
<tr>
<td>Bio15</td>
<td>Precipitation seasonality (coefficient of variation)</td>
</tr>
<tr>
<td>Bio16</td>
<td>Precipitation of wettest quarter</td>
</tr>
<tr>
<td>Bio17</td>
<td>Precipitation of driest quarter</td>
</tr>
<tr>
<td>Bio18</td>
<td>Precipitation of warmest quarter</td>
</tr>
<tr>
<td>Bio19</td>
<td>Precipitation of coldest quarter</td>
</tr>
<tr>
<td>DEM</td>
<td>Digital elevation model</td>
</tr>
<tr>
<td>Slope</td>
<td>Slope</td>
</tr>
<tr>
<td>Flow direction</td>
<td>Direction of stream drainages</td>
</tr>
<tr>
<td>TCC</td>
<td>Tree canopy coverage</td>
</tr>
</tbody>
</table>