

ORIGINAL RESEARCH

Phylogeography and population genetics of pine butterflies: Sky islands increase genetic divergence

Dale A. Halbritter¹  | Caroline G. Storer² | Akito Y. Kawahara² | Jaret C. Daniels^{1,2}

¹Entomology and Nematology Department, University of Florida, Gainesville, FL, USA

²McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, University of Florida, Gainesville, FL, USA

Correspondence

Dale A. Halbritter, Entomology and Nematology Department, Steinmetz Hall, University of Florida, 1881 Natural Area Drive, Gainesville, FL 32611, USA.
Email: dhalb001@ufl.edu

Present address

Dale A. Halbritter, USDA-ARS Invasive Plant Research Laboratory, 3225 College Ave, Fort Lauderdale, FL 33314, USA

Funding information

McGuire Center for Lepidoptera and Biodiversity; Division of Environmental Biology, Grant/Award Number: DEB-1541500; University of Florida; American Museum of Natural History

Abstract

The sky islands of southeastern Arizona (AZ) mark a major transition zone between tropical and temperate biota and are considered a neglected biodiversity hotspot. Dispersal ability and host plant specificity are thought to impact the history and diversity of insect populations across the sky islands. We aimed to investigate the population structure and phylogeography of two pine-feeding pierid butterflies, the pine white (*Neophasia menapia*) and the Mexican pine white (*Neophasia terlooii*), restricted to these “islands” at this transition zone. Given their dependence on pines as the larval hosts, we hypothesized that habitat connectivity affects population structure and is at least in part responsible for their allopatry. We sampled DNA from freshly collected butterflies from 17 sites in the sky islands and adjacent high-elevation habitats and sequenced these samples using ddRADSeq. Up to 15,399 SNPs were discovered and analyzed in population genetic and phylogenetic contexts with Stacks and pyRAD pipelines. Low genetic differentiation in *N. menapia* suggests that it is panmictic. Conversely, there is strong evidence for population structure within *N. terlooii*. Each sky island likely contains a population of *N. terlooii*, and clustering is hierarchical, with populations on proximal mountains being more related to each other. The *N. menapia* habitat, which is largely contiguous, facilitates panmixia, while the *N. terlooii* habitat, restricted to the higher elevations on each sky island, creates distinct population structure. Phylogenetic results corroborate those from population genetic analyses. The historical climate-driven fluxes in forest habitat connectivity have implications for understanding the biodiversity of fragmented habitats.

KEYWORDS

ddRADSeq, *Neophasia*, phylogeny, Pieridae, *Pinus*, population

1 | INTRODUCTION

The isolated mountain habitats of western North America provide a natural laboratory for investigating the evolutionary processes at work in taxa restricted to naturally or anthropogenically fragmented habitats. The Madrean Archipelago (sky islands) is a region of

isolated mountain ranges spanning extreme southeastern Arizona, USA (AZ), southwestern New Mexico, USA, and northern Mexico. The lower elevations (<1,000 m asl) make up the Sonoran Desert to the west and the Chihuahuan Desert to the east. The higher elevations (>2,000 m asl) are interspersed in the gap between the Rocky Mountains to the north and the Sierra Madre Occidental to the south

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

(Figure 1). The arid desert lowlands can serve as a barrier to dispersal for high-elevation animal species (Holycross & Douglas, 2007; Masta, 2000; Ober, Matthews, Ferrieri, & Kuhn, 2011; Tennessen & Zamudio, 2008). The sky islands are also a region of high species turnover, as dozens of insect genera and hundreds to thousands of insect species reach the northern or southern limits of their ranges there (Felger & Wilson, 1994). As a result of the complex topography, mixing of temperate and tropical climate zones, and convergence of two major mountain ranges, the sky islands are one of the most biologically diverse ecoregions in the world (Skroch, 2008). By studying abundant and ecologically diverse taxa such as insects, we can gain a better understanding of diversification and community structure (Moore et al., 2013).

Neophasia butterflies are an excellent system for investigating how the sky islands have shaped species history. These butterflies are coniferous forest specialists and their evolutionary history likely follows that of their forest habitats, which have been hypothesized to fluctuate in area and connectivity in association with climate changes (Thompson & Anderson, 2000). The near-contact zone of ranges of the two *Neophasia* species meets at the major transition of ecoregions marked by the sky islands in AZ. *Neophasia menapia* (Felder & Felder, 1859) ranges from southwestern British Columbia to the Mogollon Rim and White Mountains of AZ (Scott, 1986), and southeastward through New Mexico to Guadalupe Mountains National Park in Texas (Lotts & Naberhaus, 2016). *Neophasia menapia* has four geographically defined subspecies (Pelham, 2008)

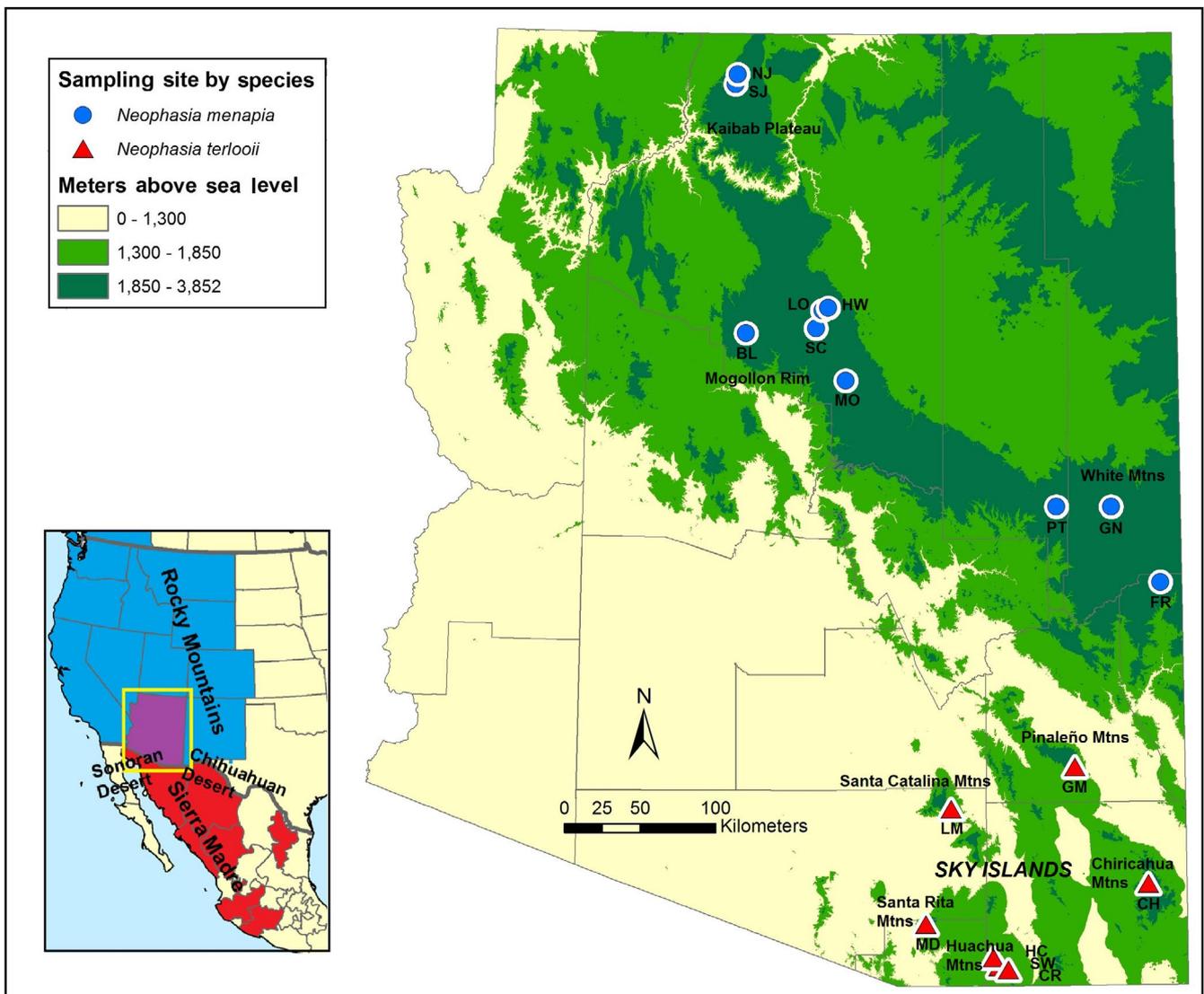


FIGURE 1 Map of Arizona, showing sampling sites and their two-letter identifiers. Elevation data were acquired from the National Resources Conservation Service Geospatial Data Gateway. The 1,300–1,850 m elevation range represents a scenario 18,000 C¹⁴ years before present when the open conifer woodlands occurred in that range, evidenced by packrat middens containing plant matter from open conifer woodlands (Table 3 of Thompson and Anderson (2000)). Elevations above 1,850 m represent the present day start of open conifer woodlands. Inset map shows major ecoregions in black text and shades the state/province-scale distributions (except for Texas, which has only one record in the extreme SW corner) of each *Neophasia* spp.: blue for *Neophasia menapia* and red for *Neophasia terlooii*. Both species occur in Arizona, and it is therefore shaded purple

and *N. m. magnamenapia* is the subspecies in AZ on which this study was focused (henceforth *N. menapia*). *Neophasia terlooii* (Behr, 1869) is present in the sky islands of southeastern AZ southward to the Sierra Madre Occidental in central Mexico (Bailowitz & Brock, 1991). Both *N. menapia* and *N. terlooii* are exemplary representatives of the two major ecoregions that meet at the sky islands: *N. menapia* from the Rocky Mountain ecoregion and *N. terlooii* from the Madrean ecoregion. Additionally, because *Neophasia* are locally abundant, easy to spot, collect, and identify, and are fairly weak flyers, they are reliably available for collection in sufficient numbers at multiple sites. Additionally, due to their dependence on climatologically sensitive pine forests, they constitute an excellent model taxon to investigate how climate-induced changes in habitat connectivity influence diversification.

The larval hosts of *N. menapia* include *Abies*, *Pseudotsuga*, *Tsuga*, *Picea*, and *Pinus*, (Cole, 1971; Evenden, 1926; Robinson, Ackery, Kitching, Beccaloni, & Hernández, 2002; Scott, 1986) and *N. terlooii* feeds only on the latter two genera (Arizona Game & Fish Department, 2001). *Pinus ponderosa* (Engelmann), a larval host for both *Neophasia* species, occurs in the more fragmented Madrean evergreen woodlands and is a predominant member of the comparably larger and more contiguous Rocky Mountain coniferous forests (Brown, 1994). There is some overlap in the voltinism of these two species in AZ; *N. menapia* is univoltine, adults are active from mid-July through August, and *N. terlooii* is bivoltine and is active in low numbers from June through early September with peak adult activity in October. For both species, eggs are laid on live pine needles. Some populations of *N. menapia* in Montana have occasionally undergone population irruptions (Dewey, Ciesla, & Meyer, 1974), but in most cases *Neophasia* are locally abundant and not believed to disperse far from forested habitats containing their host trees.

Given the close and abrupt range disjunction in *Neophasia*, our main objective was to determine whether *Neophasia* species residing in forested habitats of different ecoregions of the sky islands and surrounding area have similar concordant population structure. Using newly discovered SNPs from ddRADSeq libraries generated

for this study, we infer with multiple lines of evidence the population structure and phylogeographic history of *Neophasia*. We hypothesized that, despite their capability of flight, genetic structuring of *Neophasia* populations would reflect the degree of current habitat connectivity. We discuss the limits of *Neophasia* dispersal with respect to the geographic layout of their mountain habitats in Arizona. In addition to physiological constraints and climate differences between habitats (Halbritter, Teets, Williams, & Daniels, 2018), the desert and grassland matrices separating the two species' mountain habitats are likely to be dispersal barriers contributing to allopatry.

2 | MATERIALS AND METHODS

2.1 | Specimens

Neophasia menapia ($n = 88$) and *N. terlooii* ($n = 69$) were collected from 10 and seven field sites, respectively, throughout AZ in the summer and fall of 2013 and 2014 (Figure 1, Table 1). Specimens were kept alive until they were stored at -15°C for a minimum of 12 hr. Frozen butterflies were then placed into 95% ethanol stored at the same temperature until they were moved to the University of Florida, Gainesville, FL, USA, at which point they were stored at -80°C .

2.2 | Library preparation

Thoracic muscle tissue, roughly $3\text{ mm} \times 1\text{ mm}$, was removed from each adult butterfly for DNA extraction. Genomic DNA was extracted from tissues using the OmniPrep extraction kit (G-Biosciences; Cat. # 786-136) following the manufacturer's protocol, with some adjustments to optimize DNA recovery (Appendix S1). Each DNA extract was quantified using a Qubit dsDNA BR Assay Kit (Life Technologies, Co.; Lot # 1681328). Extractions that contained a DNA concentration of at least $10\text{ ng}/\mu\text{l}$ were used for library preparation.

Library preparation for ddRADSeq followed that of Peterson, Weber, Kay, Fisher, and Hoekstra (2012), with some added quality control measures (Appendix S1). We used the restriction enzymes EcoRI (NEB, R3101 20,000 units/ml) and MseI (NEB, R0525 10,000 units/ml). These two enzymes have been successfully used to obtain 20,737 SNPs in an independent population genomics study on *N. menapia* (Bell, 2012). Unique barcodes, ranging from 8 to 14 bases in length, and adaptors were then ligated to the EcoRI ends of digestion fragments and common adaptors were ligated to the MseI termini. Each barcoded individual was PCR amplified using Illumina sequencing primers to confirm the success of digestion and ligation reactions. Individuals were then pooled into a single sequencing library. Enzymes and buffers were removed from the pooled library using the QIAquick PCR Purification Kit (Qiagen; Cat. # 28106), fractionated, and fractions between 275 and 475 bases were collected using the Pippin ELF (Sage Science) platform. Illumina flowcell binding sequences were incorporated into the desired fractionation product using PCR, primers were removed using AMPure beads (Beckman Coulter Life Sciences), and lastly single-end DNA

TABLE 1 Number of butterflies per sampling site for the ddRADSeq analyses, with ten being the target number

Species	Site	<i>n</i>	Species	Site	<i>n</i>
<i>Neophasia menapia</i> ^a	BL	10	<i>Neophasia terlooii</i>	CH	10
	FR	10		CR	10
	GN	10		GM	10
	HW	3		HC	9
	LO	6		LM	10
	MO	8		MD	10
	NJ	9		SW	10
	PT	6			
	SC	10			
	SJ	10			

^aEighty-eight *N. menapia* were initially collected, but six were excluded after data filtering.

sequencing was performed on the Illumina NextSeq500 platform using a midoutput flowcell for 150 cycles at the University of Florida Interdisciplinary Center for Biotechnology Research.

2.3 | Raw data processing

Raw sequences containing the multiplexed reads (10.7 GB total) were first assessed for quality by generating FastQC reports (FastQC 0.11.4; Andrews, 2010). Any reads that did not have at least 90% of bases with a Phred score of ≥ 20 were discarded using FastX quality trimmer from the FASTX_Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Filtered sequences were then demultiplexed using FastX_Toolkit FASTQ Barcode splitter and either the 14-, 10-, 9-, or 8-base barcodes. FastQC reports, filtering, and demultiplexing were all performed using the UF instance of Galaxy (Afgan et al., 2018).

Barcodes and the 6-base pair enzyme cut site were trimmed from the 5' end and 5–10 bases were trimmed from the 3' end of the reads from each individual. Bases were trimmed from the 3' end because the last 5–10 bases tended to have lower Phred scores. Reads were all trimmed to be at most 125 bases in length to facilitate downstream analyses in Stacks. Trimming was accomplished using FASTX_trimmer from the FASTX_Toolkit. Because we used two restriction enzymes, some of the MseI cut sites were less than 125 bases from the EcoRI cut site at the 5' end. These shorter reads were removed using FASTX_clipper to yield the final library of demultiplexed files for each individual.

2.4 | Discovering loci and calling genotypes

We used STACKS 1.35 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) for locus discovery and genotyping. Because a reference genome was not available to assemble loci when processing the reads, we ran the STACKS program `denovo_map.pl` to assemble loci. Unless otherwise specified with a population map, STACKS assumes all barcoded individuals belong to a single population. To avoid investigator bias during SNP identification, we did not specify a population map because we did not know if our sampling sites could be considered true populations. Key parameters were set within the range to optimize the discovery of loci and SNPs: `-m 3 -M 3 -n 2 -T 6 -t -S -b 1` (Mastretta-Yanes et al., 2015). To produce genepop files and generate summary statistics, the STACKS outputs from `denovo_map.pl` were run through the populations component of STACKS with the following parameters: `-b 1 -k -p 1 -m 7 -r 0.80 -f p_value -t 8 --structure --write_single_snp --genepop --fasta`. The populations component of STACKS was run on each species separately to maximize the number of loci retained within each species. This was accomplished by specifying population maps that included only *N. menapia* as one population or *N. terlooii* as one population. For both species, we omitted from downstream analyses any individuals with <60% of genotypes present. This was a conservative filtering step given that Mastretta-Yanes et al. (2015) used a cutoff of 50%. The STACKS populations pipeline was then run again under the same parameters

but with the filtered datasets. After such filtering, we retained 82 *N. menapia* and 69 *N. terlooii* (Table 1).

2.5 | Quantifying genetic variability and spatial relationships

For all downstream analyses with the STACKS populations outputs, R 3.2.2 was used (R Core Team, 2015). Expected heterozygosity was computed using function *Hs* from “ade4genet” (Jombart, 2008) and deviations from Hardy–Weinberg equilibrium were tested for using the *hw.test* function from “pegas” (Paradis, 2010). We did not know if our sampling sites represented true populations; therefore, we first quantified genetic differentiation by comparing fixation indices (F_{ST}) between sampling locations within each species. To determine between location fixation indices, we ran *pairwise.fst* from “heirfstat” (Goudat & Jombart, 2015) for populations within each species and computed F_{ST} confidence intervals using *boot.ppfst*. Our sampling was spatially clustered with respect to site; therefore, F_{ST} comparisons were appropriate (Manel, Schwartz, Luikart, & Taberlet, 2003). A test for isolation by distance was run for each species using *mantel.randtest* from “ade4” (Dray & Dufour, 2007). An isolation by distance plot was generated, and F_{ST} values were used to plot genetic distance as a function of geographic distance.

An analysis of molecular variance (AMOVA) was performed on the STACKS-generated genotypes for both species using *poppr.amova* from “poppr” 2.1.1 (Kamvar, Tabima, & Grünwald, 2014). Strata were specified such that each individual was placed in a sampling region (mountain range) and then a site within each region (e.g., there were three sampling sites within the Huachuca Mountains). To test the significance of variations within individuals, between individuals, between sites (i.e., collection sites), and between mountains (i.e., mountain ranges or regions containing one or more sites), *randtest* from “ade4” was used on the AMOVA objects.

2.6 | Identifying population clusters

To identify genetic clusters and estimate cluster membership probabilities of each individual, we used STRUCTURE 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) to estimate *K*, searching from *K* = 1–15 with 10 replicates per *K* and a burnin of 100,000 and 150,000 reps after burnin for each replicate. Prior population or location information was not used for inferring clusters and admixture was allowed. Results were interpreted using Structure Harvester (Earl & vonHoldt, 2012) and plots generated using Structure Plot (Ramasamy, Ramasamy, Bindroo, & Naik, 2014). Substructure was investigated in *N. terlooii* for sampling sites in which all individuals had at or near 100% probability assignment to one of two clusters (i.e., minimal admixture). Two datasets, each including only minimally admixed sampling sites from one of the clusters, were analyzed again with STRUCTURE using the aforementioned parameters, except that only *K* = 1–5 were tested.

2.7 | Phylogenetic analysis

To further investigate genetic structure, we ran the ddRADSeq data through the *pyRAD* pipeline to infer phylogenetic relationships. Eaton (2014) developed *pyRAD* to maximize phylogenetic information across more distantly related samples from RADseq data. *STACKS* and *pyRAD* differ in that *pyRAD* uses a global alignment algorithm that allows indels. Although the *STACKS* pipeline has been used to generate species-level phylogenies (e.g., Jones, Fan, Franchini, Schartl, & Meyer, 2013), *pyRAD* can be equally as effective and generally discovers more loci than *STACKS* (Pante et al., 2015). For example, Mort et al. (2015) utilized the *pyRAD* pipeline to generate a well-resolved phylogeny of a plant genus.

In *pyRAD*, adaptor-trimmed sequences with quality scores were used for demultiplexing. As a quality control measure, reads were kept if there were fewer than 15 sites that had Phred scores ≤ 20 . The minimum number of reads for cluster formation and putative loci identification was set to seven. Putative *pyRAD* loci were retained if at least ten individuals shared a locus. If we required more individuals to share a locus, we lost a considerable number of loci (see Results); therefore, we kept the minimum number of individuals (i.e., samples in *pyRAD*) to ten. Up to 100 SNPs per locus (default for parameter `## 26.opt`) were allowed. Using *pyRAD* generated data matrices, maximum likelihood trees with 10,000 ultrafast bootstrap replicates (Minh, Nguyen, & Haeseler, 2013) were inferred using IQ-TREE (Nguyen, Schmidt, Haeseler, & Minh, 2015). The model used for both species was GTR + I + G4. Tree files were visualized using FIGTREE (Rambaut, 2007).

TABLE 2 A comparison of the number of loci discovered for either *Neophasia* spp. with the *pyRAD* and *STACKS* pipelines

Species	Number of loci per pipeline	
	<i>pyRAD</i>	<i>STACKS</i>
<i>Neophasia menapia</i>	10,740	4,732
<i>Neophasia terlooii</i>	9,351	3,125

The same individuals were included in each pipeline.

TABLE 3 Heterozygosity estimates for each geographic sampling site

Species	Site	Heterozygosity estimate	Species	Site	Heterozygosity estimate
<i>Neophasia menapia</i>	BL	0.1316077	<i>Neophasia terlooii</i>	CH	0.16146579
	FR	0.1272531		CR	0.14976759
	GN	0.1297881		GM	0.09822872
	HW	0.1306090		HC	0.16208089
	LO	0.1276648		LM	0.12387499
	MO	0.1307747		MD	0.14302912
	NJ	0.1318904		SW	0.16317188
	PT	0.1224726			
	SC	0.1326815			
	SJ	0.1336801			

3 | RESULTS

A cumulative total of approximately 120 million raw reads were generated from 151 individuals across ten *N. menapia* and seven *N. terlooii* sampling sites in AZ. All of the *N. terlooii* individuals were retained after filtering and 82 of the 88 *N. menapia* were retained. After filtering, up to 10,740 were loci were recovered per sampling site, depending on the method and species (Table 2). Each locus contained one SNP for *STACKS* and up to 100 SNPs for *pyRAD*.

3.1 | Quantifying genetic variability and spatial relationships

Within the filtered individuals, 90% of *N. menapia* loci and 70.8% of *N. terlooii* loci were in Hardy–Weinberg equilibrium. Heterozygosity estimates were greater within *N. terlooii* sampling locations than within *N. menapia* sampling locations (Table 3). Fixation indices between *N. terlooii* sites were, on average, five times greater than those between *N. menapia* sites: *N. menapia* site pairs ranged from $7.31e^{-5}$ to 0.061, while *N. terlooii* site pairs ranged from 0.034 to 0.267 (Table S1). There were significant amounts of variation between mountains and between sampling sites within mountains for both *N. menapia* and *N. terlooii*, and there was a significant amount of variation within *N. terlooii* individuals as indicated by the analysis of molecular variance (Table 4). When plotting genetic distance as a function of geographic distance, there was a positive correlation between the two for *N. terlooii* (Figure 2b). This was evident with significant isolation by distance ($r = .9062$, $p = .001$), from a Mantel test. *Neophasia menapia* did not show significant isolation by distance ($r = -.081$, $p = .583$), nor did genetic distance correlate with geographic distance (Figure 2a).

3.2 | Identifying population clusters

For *N. terlooii*, Bayesian hierarchical clustering using *STRUCTURE* suggested an optimum *K* of 2 based on delta *K* (Evanno, Regnaut, & Goudet, 2005) and the mean log probabilities (Table 5). Substructure was evident in one of the primary clusters in which the site in the

Santa Rita Mountains separated from the three sites in the Huachuca Mountains. An optimal K for *N. menapia* was not clear using these methods with ΔK suggesting a $K = 9$, but mean log probabilities increasing with K and the highest probability being for the maximum K tested: $K = 15$ (Table 5). Assignment probability plots of individuals to each of several population clustering options are presented in Figure 3.

3.3 | Phylogenetic analysis

pyRAD allowed us to use the ddRADSeq data for inferring phylogenetic relationships among individuals across our sampling sites. Specifying that 10% of individuals must share a locus produced large matrices with a lot of missing data in the analysis. However, using a more stringent threshold, such as requiring a minimum of 80% of individuals to share a locus (as used in the STACKS populations pipeline for genotyping) resulted in unresolved trees with low branch support. Despite the risk of greater missing data, which can negatively affect accuracy in smaller datasets (Lemmon, Brown, Stanger-Hall, & Lemmon, 2009), gathering a greater number of loci and generating larger supermatrices results in more fully resolved phylogenies (Wagner et al., 2013). Unrooted trees were used to depict the relationships between individuals. *Neophasia menapia* formed a polytomy, but there were a few well-supported clades of up to three individuals from the same sampling sites (Figure 4a). Individuals of *N. terlooii* formed well-supported clades corresponding to each mountain range (Figure 4b). Individuals from sites within the Huachuca Mountains were polyphyletic.

4 | DISCUSSION

4.1 | Overview

The sky islands are considered a neglected biodiversity hotspot (Felger & Wilson, 1994). Research on the biodiversity of the sky

islands has focused primarily on dispersal and vicariance events associated with the Pleistocene epoch with subsequent divergence associated with biotic (Carstens & Knowles, 2007; Tennessen & Zamudio, 2008) and abiotic factors (Smith & Farrell, 2005). Within the sky islands, geographic structuring in genetic relationships has been observed in flightless beetles (Ober et al., 2011; Smith & Farrell, 2005), an ant (Favé et al., 2015), and a jumping spider (Masta, 2000). Our work here demonstrates that, despite their flight capability, *N. terlooii* has highly structured genetic patterns, which may be the consequence of abiotic factors, such as climate change and spatial habitat connectivity. The AZ sky islands and surrounding areas occupied by *Neophasia* are natural landscape-scale laboratories, providing comparisons for how connectivity influences the evolution of organisms in fragmented habitats.

Our results support the hypothesis that habitat connectivity affects the genetic structure of *Neophasia*. Our geographic sampling strategy facilitated population genetic comparisons between individuals in large, contiguous patches and those in small, isolated patches. We were also able to compare individuals between clusters of proximal and distant sampling sites within single patches. Nearly all evidence suggests that *N. menapia* is panmictic and there is minimal genetic structure with respect to sampling sites, most of which are connected by a largely contiguous habitat belt. There is strong evidence for genetic structure within *N. terlooii* as populations of butterflies on different sky islands are distinct from each other. Despite the dispersal abilities of *Neophasia* butterflies, present day desert and grassland matrices between coniferous forests limit their dispersal.

4.2 | *Neophasia menapia* is panmictic in AZ

Traditionally, population genetic studies required at least tens of individuals per sampling site to make accurate inferences. Willing, Dreyer, and Oosterhout (2012) found that the original Wright (1951) F_{ST} overestimated differentiation with small sample sizes ($n \leq 6$), but

TABLE 4 AMOVA for each species of *Neophasia*. Some mountains contain multiple collection sites

Comparison	Df	Sum sq	Mean sq	Sigma	% Covariance	p-Value
<i>Neophasia menapia</i>						
Within samples	4	475.7429	118.93572	0.5333659	0.5519891	.2524
Between samples within site	5	511.4623	102.29247	0.4818000	0.4986227	.4676
Between site within mountain	64	6,132.8688	95.82607	0.2150782	0.2225879	.0146*
Between mountain	74	7,059.2980	95.39592	95.3959184	98.7268002	.0311*
Total	147	14,179.3720	96.45831	96.6261625	100.0000000	
<i>Neophasia terlooii</i>						
Within samples	65	4,622.5000	71.11538	71.11538462	73.2834232	<.0001***
Between samples within site	58	4,114.6847	70.94284	-0.08627229	-0.0889024	.5137
Between site within mountain	2	205.2079	102.60394	1.76621812	1.8200634	.0015**
Between mountain	4	2,755.1536	688.78839	24.24623981	24.9854158	.0001***
Total	129	11,697.5462	90.67865	97.04157026	100.0000000	

Note: Asterisks denote statistical significance.

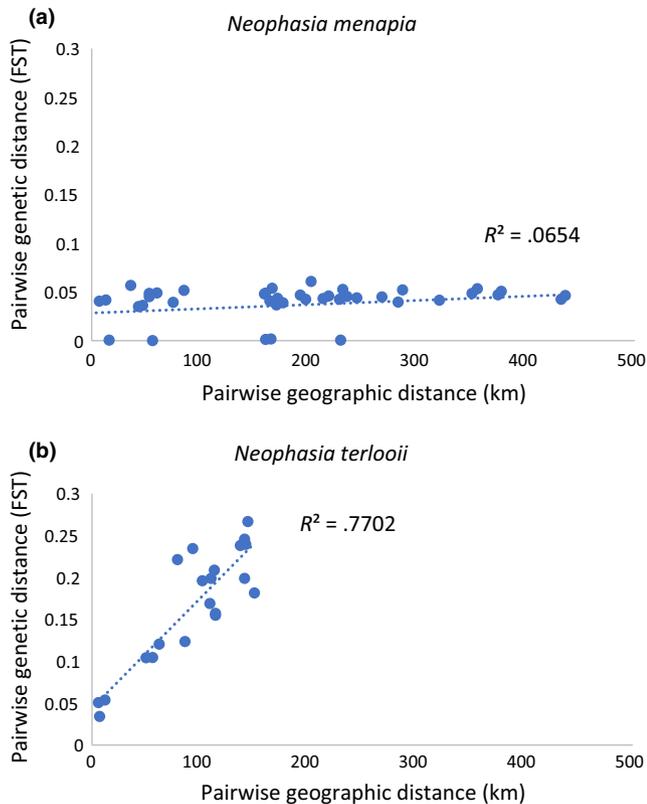


FIGURE 2 Plot of genetic distance (F_{ST}) as a function of geographic distance for (a) *Neophasia menapia* and (b) *Neophasia terlooii*. Each point represents a pairwise comparison of two individuals from different collection sites

sample sizes from individual populations can be small ($n = 6$) provided there is a large number of genetic markers ($k > 1,000$). In our study, at least 3,000 SNPs were genotyped using *STACKS*, and thus our sample sizes, averaging ten individuals per site, were appropriate for making F_{ST} comparisons.

Of the 36 pairwise F_{ST} indices for *N. menapia*, 28 were < 0.05 , suggesting panmixia. The AMOVA indicated that there was little, but significant genetic variation between sites for *N. menapia*, but additional analyses suggested this variation is not sufficient to generate robust population or phylogenetic inferences. The two *N. menapia* sites that were most distant geographically from each other (NJ and FR at 436 km) had an F_{ST} value of 0.046, but the two closest *N. terlooii* sites (HC and SW at 7.6 km) had a similar F_{ST} of 0.034 (compare Figure 1 and Table S1). Most other *N. terlooii* population pairs on different mountains had F_{ST} indices of approximately 0.20. This suggests that, despite the greater geographic distance between several *N. menapia* sampling sites, gene flow is occurring or has occurred more frequently between *N. menapia* sites than between *N. terlooii* populations.

The sampling sites for *N. menapia* were all largely on one habitat island. However, the Grand Canyon geographically isolated the sites NJ and SJ on the Kaibab Plateau from the rest of the sites to the south of the canyon. The connectivity of habitat is likely responsible for the apparent panmixia of *N. menapia*, as ponderosa pine forest is largely contiguous and encompasses all sampling sites south

of the Grand Canyon. The ddRADSeq phylogeny was essentially a polytomy (Figure 4a). There is some evidence for the populations north of the Grand Canyon to be distinct from those south of the canyon. This is apparent in the *STRUCTURE* plot (Figure 3b). However, the *STRUCTURE* analysis did not produce a clear, definitive K , which appears to provide further evidence for panmixia. We conclude that *N. menapia* is most likely panmictic in AZ and flight across the Grand Canyon is possible. However, given the possibility that some ancestral polymorphism was retained, the Grand Canyon may restrict gene flow to some degree given the lack of host plants at the lower elevations across the canyon.

Isolation by distance (IBD; Wright, 1943) occurs when genetic differentiation is due to limited gene flow across long distances rather than biologically relevant barriers in the environment. Populations arising from dispersal events tend to be geographically clustered and represent some pattern of spatial autocorrelation (Meirmans, 2012). Mantel tests cannot differentiate between patterns arising from such clustering and those from IBD (Meirmans, Goudet, & Gaggiotti, 2011). The latter authors argue that it is best to use a clustering method that accounts for the geographical locations of the samples. The Mantel test suggested that IBD was not significant between *N. menapia* sampling sites, which also did not show any consistent or well-supported clustering. This makes sense because there were roughly equivalent F_{ST} indices between proximal and distant sites (Figure 2a). Because we ran the Mantel test without specifying the geographical location data and there was evidence for hierarchical clustering within *N. terlooii* sampling sites, it is not surprising that the Mantel test was significant for *N. terlooii*. We conclude that the isolation by distance within *N. terlooii* is more likely due to hierarchical clustering on isolated mountains.

4.3 | Sky islands are responsible for *Neophasia terlooii* population structure

During our surveys, *Neophasia* butterflies were only observed at elevations between 1,700 and 2,600 m above sea level, and these elevations contain the host trees. The lowering of forest ecotones during the last glacial maxima (Thompson & Anderson, 2000) would have had little impact with respect to changes in past connectivity of our *N. menapia* sampling sites, but it would have connected several *N. terlooii* populations in the sky islands (Figure 1). The *N. terlooii* population in the Santa Rita Mountains (MD) would have been connected with the population in the Huachuca Mountains (sites HC, SW, and CR) and this largely agrees with the molecular data, as it is clustered with the latter three in the first *STRUCTURE* analysis. The population in the Pinaleño Mountains (GM) would have been connected with the population in the Chiricahua Mountains (CH), and this is evidenced by admixture with the former. The Santa Catalina Mountains (LM) is the only range that would not have been connected, yet the molecular data suggest admixture with CH and membership to a cluster also containing GM.

Neophasia terlooii populations show a very clear hierarchical population structure with respect to populations on each sky island. The

TABLE 5 Bayesian clustering results for possible numbers of clusters of individuals (*K*)

Species	<i>K</i>	Reps	Mean LnP(<i>K</i>)	Stdev LnP(<i>K</i>)	Ln'(K)	Ln''(K)	Delta <i>K</i>
<i>Neophasia menapia</i>	1	20	-157,726.23	5.28	NA	NA	NA
	2	20	-157,737.73	6.31	-11.50	13.63	2.16
	3	20	-157,735.60	4.97	2.13	16.23	3.27
	4	20	-157,717.24	7.34	18.36	3.83	0.52
	5	20	-157,702.71	3.63	14.53	12.10	3.33
	6	20	-157,700.28	4.95	2.43	0.20	0.04
	7	20	-157,697.65	3.26	2.63	7.07	2.17
	8	20	-157,687.95	4.19	9.70	13.29	3.17
	9	20	-157,691.54	2.28	-3.59	9.57	4.20
	10	20	-157,685.57	4.61	5.98	5.67	1.23
	11	20	-157,685.26	7.88	0.31	0.65	0.08
	12	20	-157,684.30	3.05	0.96	3.95	1.29
	13	20	-157,679.39	3.40	4.91	8.08	2.37
	14	20	-157,682.56	5.11	-3.17	6.74	1.32
	15	20	-157,678.99	2.76	3.57	NA	NA
<i>Neophasia terlooii</i>	1	20	-138,566.38	0.77	NA	NA	NA
	2	20	-123,727.67	9.46	14,838.72	25,799.73	2,725.93
	3	20	-134,688.68	255.10	-10,961.02	11,796.17	46.24
	4	20	-133,853.53	718.25	835.15	61.90	0.09
	5	20	-133,080.28	318.51	773.26	1,267.90	3.98
	6	20	-133,574.92	179.67	-494.64	647.31	3.60
	7	20	-133,422.25	494.84	152.67	2,053.04	4.15
	8	20	-135,322.63	1,196.98	-1,900.38	3,744.06	3.13
	9	20	-133,478.94	36.89	1,843.69	3,222.10	87.34
	10	20	-134,857.35	1,271.75	-1,378.41	1,063.56	0.84
	11	20	-135,172.20	715.26	-314.85	53.01	0.07
	12	20	-135,434.04	295.09	-261.84	963.96	3.27
	13	20	-136,659.84	129.87	-1,225.80	4,117.10	31.70
	14	20	-133,768.55	491.69	2,891.30	4,571.78	9.30
	15	20	-135,449.03	1,248.08	-1,680.49	NA	NA

Note: Bold values indicate the most likely option(s) for *K*.

northwest–southeast oriented topographic grain of the sky islands (see Felger & Wilson, 1994) could explain their north–south-nested hierarchy. Similarly, the ant *Monomorium emersoni* Gregg exhibited north and south groupings in the sky islands (Favé et al., 2015). *Neophasia terlooii* populations LM and GM are on the two northern-most sky islands and are distinct from the remaining southern populations (Figure 3b). At the next level in the hierarchical structure within the remaining southern clusters, the population in the Santa Rita Mountains (MD) is separated into its own cluster. As discussed above, MD would have been isolated historically. The sites within the Huachuca Mountains (HC, SW, and CR) form the final population with respect to mountain. Given the admixture between northern and southern clusters seen in CH, this range may have been a stepping stone between the two.

The panmixia of *N. terlooii* within the Huachuca Mountains is reflected in the ddRADSeq phylogeny where these sites are

polyphyletic (Figure 4b) and in the STRUCTURE plot where these sites are all assigned to one cluster (Figure 3b). The optimum *K* of 2 found using STRUCTURE reflects the uppermost level in the aforementioned nested hierarchy, but with some admixture between clusters for the CH population. Based on the STRUCTURE analysis and phylogeny, four populations are most likely.

4.4 | Dispersal limitations and biogeographical considerations for *Neophasia*

It is evident that the geographic distance between the sky island mountains is enough to isolate *N. terlooii* populations. The shortest distance between any two sky islands in our study is 55 km (between the Santa Rita and Huachuca Mountains, populations MD and HC + SW + CR, respectively). From the Kaibab Plateau to the south

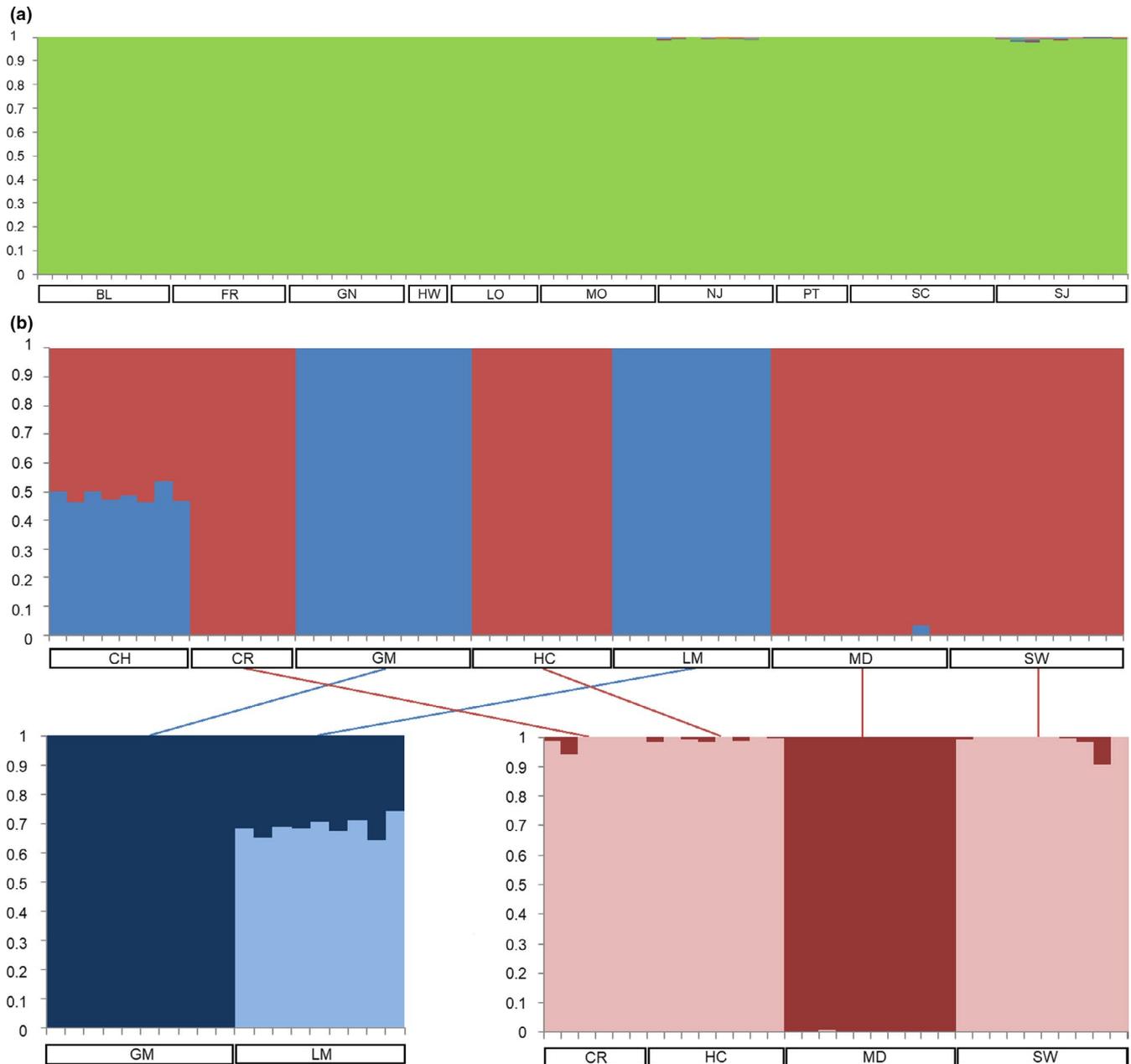


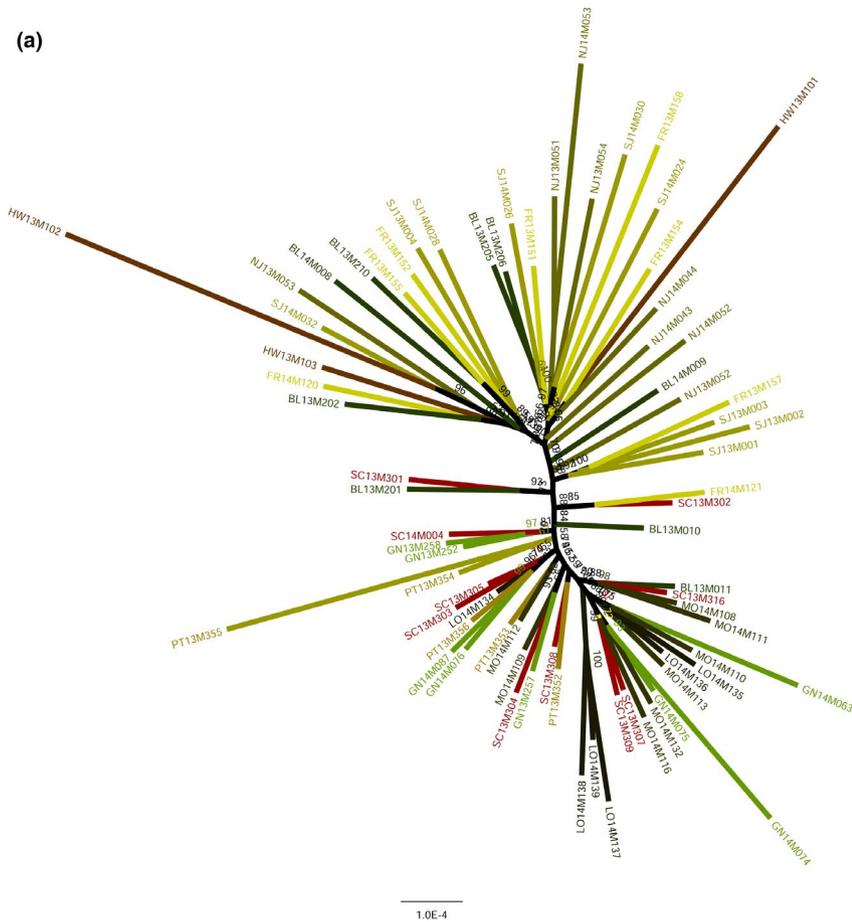
FIGURE 3 STRUCTURE clustering for (a) *Neophasia menapia* and (b) *Neophasia terlooii*. $K = 9$ is plotted for *N. menapia*. Substructure is shown in *N. terlooii* in which the cluster containing CR, HC, MD, and SW can be broken down to two clusters, one with MD and the other with CR, HC, and SW. Individuals are grouped by sampling site, indicated by the open boxes beneath each diagram

rim, the Grand Canyon spans roughly 13 km of unsuitable habitat for *N. menapia* and there not strong evidence to suggest this distance inhibits gene flow. At their closest point, the northern-most sky island and ponderosa pine habitat in the southern White Mountains are roughly 70 km apart. This distance appears sufficient to maintain allopatry between *N. menapia* and *N. terlooii*, but geographic distance may not be the only factor.

Even if the geographic distance is within the dispersal capabilities of an organism, or if the habitat between populations is suitable, allopatry can persist for biological reasons such as mating incompatibility, competition, or predation (Tennissen & Zamudio, 2008). The female *N. terlooii* has long

been suspected a mimic of the monarch butterfly (Poulton, 1914), and, as discussed in Halbritter, Gordon, Keacher, Avery, and Daniels (2018), mimicry may be advantageous to *N. terlooii* while *N. menapia* may be vulnerable to predation if it were to venture into the range of *N. terlooii*. To date, *N. menapia* has not been collected or seen in the sky islands and *N. terlooii* has not been collected or seen north of the sky islands. Crossing the barrier between the sky islands and the White Mountains is likely a rare event, and any transplants are incapable of starting new populations for reasons that are beyond the scope of this paper. Further sampling of the genetic diversity of both species of *Neophasia* throughout their respective

(a)



(b)

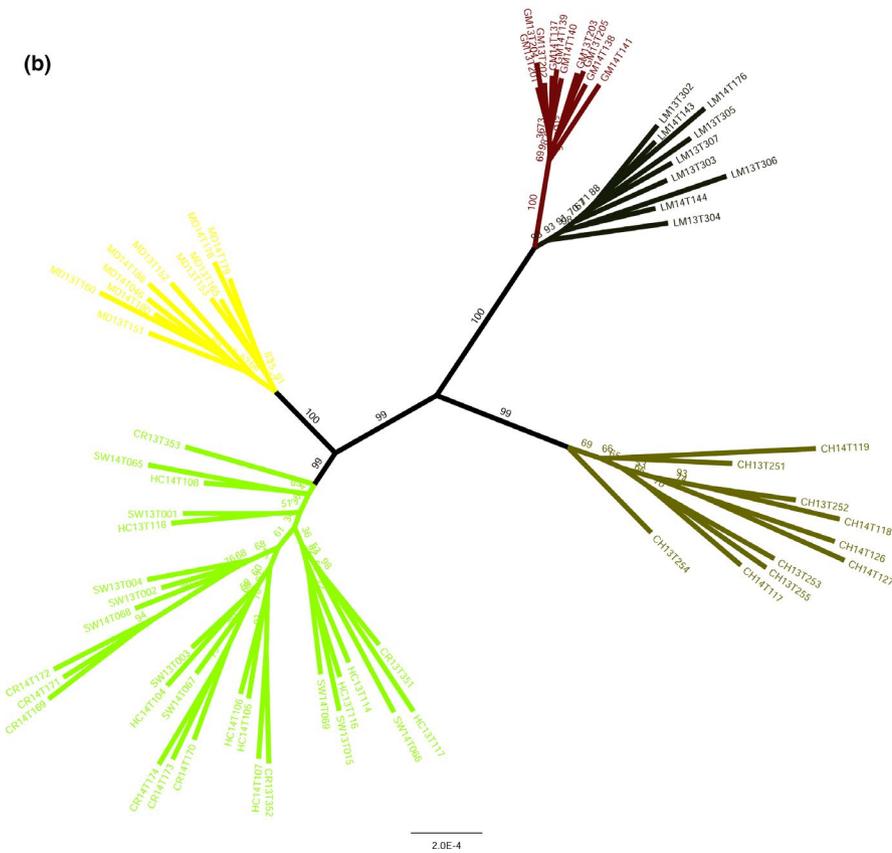


FIGURE 4 Maximum likelihood trees from ddRADSeq data for *Neophasia* spp. from Arizona. Each tip represents an individual color coded to its sampling site. Tip labels refer to sampling site abbreviation, year collected, and specimen number. Branch labels indicate support from 10,000 ultrafast bootstrap replicates. (a) Tree for *Neophasia menapia*, (b) tree for *Neophasia terlooii*

ranges would provide the data necessary to unravel their biogeographic history in western North America and would contribute to our work here and to knowledge of the evolution of biodiversity in the sky islands. Additionally, the timing and direction of dispersal events is key to further exploring their biogeography. This remains for future study and will require more accurate methods of divergence time estimation at the population level.

5 | CONCLUSIONS

We extracted genome-wide data from 151 individuals and used this data to infer population structure and generate phylogenies for a nonmodel taxon in a neglected biodiversity hotspot. Despite differences in the number and type of loci discovered across the genomic pipelines used, we reached congruent biological conclusions regarding *Neophasia* population structure in AZ. As our ability to analyze genomic data catches up with our ability to generate data, complex biogeographical studies using nonmodel taxa at very shallow phylogenetic levels will be increasingly common and affordable. With improved algorithms and faster processors, total-evidence approaches are becoming increasingly feasible, especially for projects with minimal or short-term funding.

ACKNOWLEDGMENTS

We thank Andrea Lucky, Wendy Moore, and Samantha Wisely for help in project planning and discussions, Jim Brock, Aaron Chambers, Donal Hill, and Alan Yanahan for help with field collecting, the Merriam-Powell Research Station, the University of Arizona Entomology Department, and Chris Bibbs and Alan Yanahan for lodging and laboratory space while in Arizona, Matthew Standridge for handling specimen shipments, Vincent Ficarrotta and Lei Xiao for assistance with DNA extraction, Stuart McDaniel and Adam Payton for technical support, materials, and laboratory space needed to assemble ddRADSeq libraries, and Jesse Breinholt, Adam Payton, and Deborah Triant for help with bioinformatics. We extend our appreciation to Fort Huachuca and the National Forest Service and for their cooperation. Some of the infrastructure used in this study was funded by the US National Science Foundation Grant no. DEB-1541500 to A.Y.K. Funds from the William C. and Bertha M. Cornett Fellowship through the University of Florida and the Theodore Roosevelt Memorial Grant and were used to cover travel and lodging expenses. Funding for the article processing charge was provided by the McGuire Center for Lepidoptera and Biodiversity. Lastly, we thank our three anonymous reviewers for their detailed comments and suggestions during revision.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

D.A.H. led this project as a part of his dissertation research; D.A.H., A.Y.K., and J.C.D. conceived the ideas; D.A.H. collected the data; D.A.H. and C.G.S. analyzed the data; and D.A.H. led the writing.

ORCID

Dale A. Halbritter  <https://orcid.org/0000-0001-8201-1448>

DATA AVAILABILITY STATEMENT

Genetic data, code, alignments, and other output files: Dryad <https://doi.org/10.5061/dryad.4mw6m906h>

REFERENCES

- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., ... Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, 46(W1), W537–W544. <https://doi.org/10.1093/nar/gky379>
- Andrews, S. (2010). FASTQC: A quality control tool for high throughput sequence data. Retrieved from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Arizona Game and Fish Department (2001). *Neophasia terlooitii* (p. 3). Unpublished abstract compiled and edited by the Heritage Data Management System, Arizona Game and Fish Department, Phoenix, AZ.
- Bailowitz, R. A., & Brock, J. P. (1991). *Butterflies of Southeastern Arizona*. Tucson, AR: Sonoran Arthropod Studies Inc.
- Bell, K. L. (2012). *Sympatric, allochronic populations of the pine white butterfly (Neophasia menapia) are morphologically and genetically differentiated*. Master's Thesis, Texas State University - San Marcos, San Marcos, TX.
- Brown, D. E. (1994). *Biotic communities; Southwestern States; Mexico*. Salt Lake City, UT: University of Utah Press.
- Carstens, B. C., & Knowles, L. L. (2007). Shifting distributions and speciation: Species divergence during rapid climate change. *Molecular Ecology*, 16(3), 619–627. <https://doi.org/10.1111/j.1365-294X.2006.03167.x>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Cole, W. E. (1971). *Pine butterfly* (pp. 1–3). Washington, DC: U.S. Department of Agriculture, Forest Service.
- Dewey, J. E., Ciesla, W. M., & Meyer, H. E. (1974). Insect defoliation as a predisposing agent to a bark beetle outbreak in eastern Montana. *Environmental Entomology*, 3, 722–722. <https://doi.org/10.1093/ee/3.4.722>
- Dray, S., & Dufour, A. B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4), 1–20. <https://doi.org/10.18637/jss.v022.i04>
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13), 1844–1849. <https://doi.org/10.1093/bioinformatics/btu121>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A

- simulation study. *Molecular Ecology*, 14(8), 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Evenden, J. C. (1926). The pine butterfly, *Neophasia menapia* Felder. *Journal of Agricultural Research*, 33(4), 339–344.
- Favé, M. J., Johnson, R. A., Cover, S., Handschuh, S., Metscher, B. D., Müller, G. B., ... Abouheif, E. (2015). Past climate change on Sky Islands drives novelty in a core developmental network gene network and its phenotype. *BMC Evolutionary Biology*, 15, 183. <https://doi.org/10.1186/s12862-015-0448-4>
- Felger, R. S., & Wilson, M. F. (1994). Northern Sierra Madre Occidental and its Apachian outliers: A neglected center of biodiversity. *U S Forest Service General Technical Report RM*, 264, 36–51.
- Goudat, J., & Jombart, T. (2015). *hierfstat: Estimation and tests of hierarchical F-statistics, version 0.04-22*. Retrieved from <http://CRAN.R-project.org/package=hierfstat>
- Halbritter, D. A., Gordon, J. M., Keacher, K. L., Avery, M. A., & Daniels, J. C. (2018). Evaluating an alleged mimic of the monarch butterfly: *Neophasia* (Lepidoptera: Pieridae) butterflies are palatable to avian predators. *Insects*, 8(4), 150. <https://doi.org/10.3390/insects9040150>
- Halbritter, D. A., Teets, N. M., Williams, C. M., & Daniels, J. C. (2018). Differences in winter cold hardiness reflect the geographic range disjunction of *Neophasia menapia* and *Neophasia terlooii* (Lepidoptera: Pieridae). *Journal of Insect Physiology*, 107, 204–211. <https://doi.org/10.1016/j.jinsphys.2018.03.005>
- Hannon, G. J. (2010). *FASTX Toolkit*. Retrieved from http://hannonlab.cshl.edu/fastx_toolkit/index.html
- Holycross, A. T., & Douglas, M. E. (2007). Geographic isolation, genetic divergence, and ecological non-exchangeability define ESUs in a threatened sky-island rattlesnake. *Biological Conservation*, 134(1), 142–154. <https://doi.org/10.1016/j.biocon.2006.07.020>
- Jombart, T. (2008). *adeigen: An R package for the multivariate analysis of genetic markers*. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jones, J. C., Fan, S. H., Franchini, P., Schartl, M., & Meyer, A. (2013). The evolutionary history of *Xiphophorus* fish and their sexually selected sword: A genome-wide approach using restriction site-associated DNA sequencing. *Molecular Ecology*, 22(11), 2986–3001. <https://doi.org/10.1111/mec.12269>
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281. <https://doi.org/10.7717/peerj.281>
- Lemmon, A. R., Brown, J. M., Stanger-Hall, K., & Lemmon, E. M. (2009). The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Systematic Biology*, 58(1), 130–145. <https://doi.org/10.1093/sysbio/syp017>
- Lotts, K., & Naberhaus, T. (2016). *Butterflies and moths of North America*. Retrieved from <http://www.butterfliesandmoths.org/>
- Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, 18(4), 189–197. [https://doi.org/10.1016/S0169-5347\(03\)00008-9](https://doi.org/10.1016/S0169-5347(03)00008-9)
- Masta, S. (2000). Phylogeography of the jumping spider *Habronatus pugillius* (Araneae: Salticidae): Recent vicariance of Sky Island populations? *Evolution*, 54(5), 1699–1711. <https://doi.org/10.1111/j.0014-3820.2000.tb00714.x>
- Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Pinero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15(1), 28–41. <https://doi.org/10.1111/1755-0998.12291>
- Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, 21(12), 2839–2846. <https://doi.org/10.1111/j.1365-294X.2012.05578.x>
- Meirmans, P. G., Goudet, J., IntraBioDiv Consortium, & Gaggiotti, O. E. (2011). Ecology and life history affect different aspects of the population structure of 27 high-alpine plants. *Molecular Ecology*, 20(15), 3144–3155. <https://doi.org/10.1111/j.1365-294X.2011.05164.x>
- Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, 30(15), 1188–1195. <https://doi.org/10.1093/molbev/mst024>
- Moore, W., Meyer, M., Eble, J. A., Franklin, K., Wiens, J. F., & Brusca, R. C. (2013). Introduction to the Arizona Sky Island Arthropod Project (ASAP): Systematics, biogeography, ecology and population genetics of arthropods of the Madrean Sky Islands. In G. J. Gottfried, P. F. Ffolliott, B. S. Gebow & L. G. Eskew (compilers), *Merging science and management in a rapidly changing world: biodiversity and management of the Madrean Archipelago III. 2012 May 1–5, Tucson, AZ* (pp. 144–168). Proceedings RMRS-P-67. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.
- Mort, M. E., Crawford, D. J., Kelly, J. K., Santos-Guerra, A., Menezes de Sequeira, M., Moura, M., & Caujapé-Castells, J. (2015). Multiplex-shotgun-genotyping data resolve phylogeny within a very recently derived insular lineage. *The American Journal of Botany*, 102(4), 634–641. <https://doi.org/10.3732/ajb.1400551>
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Ober, K. A., Matthews, B., Ferrieri, A., & Kuhn, S. (2011). The evolution and age of populations of *Scaphinotus petersi* Roeschke on Arizona Sky Islands (Coleoptera, Carabidae, Cychrini). *ZooKeys*, 147, 183–197. <https://doi.org/10.3897/zookeys.147.2024>
- Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S. C., Boisselier, M. C., & Samadi, S. (2015). Use of RAD sequencing for delimiting species. *Heredity*, 114, 450–459. <https://doi.org/10.1038/hdy.2014.105>
- Paradis, E. (2010). pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26, 419–420. <https://doi.org/10.1093/bioinformatics/btp696>
- Pelham, J. P. (2008). Catalogue of the butterflies of the United States and Canada. *Journal of Research on the Lepidoptera*, 40, 672.
- 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. <https://doi.org/10.1371/journal.pone.0037135>
- Poulton, E. B. (1914). Mimicry in North American butterflies: A reply. *Proceedings of the Academy of Natural Sciences of Philadelphia*, 66, 161–195.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.
- R Core Team (2015). *R: a language and environment for statistical computing, version 3.2.2*. Vienna, Austria: R Foundation for Statistical Computing.
- Ramasamy, R. K., Ramasamy, S., Bindroo, B. B., & Naik, V. G. (2014). STRUCTURE PLOT: A program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus*, 3, 432. <https://doi.org/10.1186/2193-1801-3-431>
- Rambaut, A. (2007). *FigTree, version 1.4.3*. Retrieved from <http://tree.bio.ed.ac.uk/software/figtree/>
- Robinson, G. S., Ackery, P. R., Kitching, I. J., Beccaloni, G. W., & Hernández, L. M. (2002). *Hostplants of the moth and butterfly caterpillars of America north of Mexico*. Gainesville, FL: The American Entomological Institute.
- Scott, J. A. (1986). *The butterflies of North America: A natural history and field guide*. Stanford, CA: Stanford University Press.
- Skroch, M. (2008). Sky islands of North America a globally unique and threatened inland archipelago. *Terrain.org: A Journal of the Built and*

- Natural Environments*, 21. <https://www.terrain.org/articles/21/skroch.htm>
- Smith, C. I., & Farrell, B. D. (2005). Phylogeography of the longhorn cactus beetle *Moneilema appressum* LeConte (Coleoptera: Cerambycidae): Was the differentiation of the Madrean sky islands driven by Pleistocene climate changes? *Molecular Ecology*, 14(10), 3049–3065. <https://doi.org/10.1111/j.1365-294X.2005.02647.x>
- Tennessen, J. A., & Zamudio, K. R. (2008). Genetic differentiation among mountain island populations of the striped plateau lizard, *Sceloporus virgatus* (Squamata: Phrynosomatidae). *Copeia*, 106(3), 558–564. <https://doi.org/10.1643/CG-06-038>
- Thompson, R. S., & Anderson, K. H. (2000). Biomes of western North America at 18,000, 6000 and 0 C-14 yr BP reconstructed from pollen and packrat midden data. *Journal of Biogeography*, 27(3), 555–584. <https://doi.org/10.1046/j.1365-2699.2000.00427.x>
- Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., ... Seehausen, O. (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, 22(3), 787–798. <https://doi.org/10.1111/mec.12023>
- Willing, E. M., Dreyer, C., & van Oosterhout, C. (2012). Estimates of genetic differentiation measured by F-ST do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE*, 7(8), e42649. <https://doi.org/10.1371/journal.pone.0042649>
- Wright, S. (1943). Isolation by distance. *Genetics*, 28(2), 114–138.
- Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, 15(1), 323–354. <https://doi.org/10.1111/j.1469-1809.1949.tb02451.x>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Halbritter DA, Storer CG, Kawahara AY, Daniels JC. Phylogeography and population genetics of pine butterflies: Sky islands increase genetic divergence. *Ecol Evol*. 2019;9:13389–13401. <https://doi.org/10.1002/ece3.5793>