Title

Regional population differentiation in the morphologically diverse, elevationally widespread Nearctic skipper Polites sabuleti

Journal

Journal of Biogeography, 42(9)

ISSN

0305-0270

Authors

Jahner, Joshua P
Forister, Matthew L
Nice, Chris C
et al.

Publication Date

2015-09-01

DOI

10.1111/jbi.12557

Peer reviewed

Journal of Biogeography (J. Biogeogr.) (2015)

ORIGINAL ARTICLE

Regional population differentiation in the morphologically diverse, elevationally widespread Nearctic skipper *Polites sabuleti*

Joshua P. Jahner[1]*, Matthew L. Forister[1], Chris C. Nice[2], James A. Fordyce[3], Joseph S. Wilson[4], Dennis D. Murphy[1], Zachary H. Marion[3] and Arthur M. Shapiro[5]

[1]*Department of Biology, Program in Ecology, Evolution, and Conservation Biology, University of Nevada, Reno, NV 89557, USA,* [2]*Department of Biology, Program in Population and Conservation Biology, Texas State University, San Marcos, TX 78666, USA,* [3]*Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA,* [4]*Department of Biology, Utah State University, Tooele, UT 84074, USA,* [5]*Center for Population Biology, University of California, Davis, CA 95616, USA*

*Correspondence: Joshua P. Jahner, Program in Ecology, Evolution, and Conservation Biology, Department of Biology, University of Nevada, Reno, NV 89557, USA.
E-mail: jpjahner@gmail.com

ABSTRACT

Aim To evaluate three phylogeographical models associated with the presence of mountains (montane vicariance, sky island and parapatry) as drivers of intraspecific diversification in the sandhill skipper, *Polites sabuleti* (Lepidoptera: Hesperiidae), a morphologically variable species found in a variety of habitats from sea level to the alpine zone.

Location Western North America.

Methods The mitochondrial cytochrome *c* oxidase subunit II region was sequenced in 189 *P. sabuleti* individuals. Mitochondrial sequences were used in a spatial analysis of molecular variance (SAMOVA) to evaluate geographical population structure. AFLP markers were also generated for 347 individuals in order to estimate admixture proportions and examine population differentiation based on the nuclear genome.

Results Twenty-five mitochondrial haplotypes and 42 anonymous AFLP loci were found across 36 collection localities. Mitochondrial variation suggests a degree of regional population structure, although at least one of the inferred population groups extends over nearly the entire geographical range of the species. Analyses of nuclear data (AFLPs) identified five genetic clusters, including one restricted to high elevations in the Sierra Nevada.

Main conclusions The distribution of genetic variation within *Polites sabuleti*, a species with a broad elevational range, does not strictly support either mountain-associated vicariance or 'sky island' isolation as the dominant process. Instead, we find complex population structure, including evidence for divergence between high- and low-elevation populations in the Sierra Nevada mountains.

Keywords
AFLP, Hesperiidae, mitochondrial DNA, mountain uplift, Pleistocene, population structure, sandhill skipper, SAMOVA, Sierra Nevada.

INTRODUCTION

At least two major geological events have driven the diversification of many animal and plant species in western North America: mountain uplift and Pleistocene glaciation (Hewitt, 2000, 2004; Avise, 2004; Shafer *et al.*, 2010; Wilson & Pitts, 2010a). A growing consensus suggests that intra- and interspecific divergence was promoted throughout the arid portions of this region by Neogene and Pleistocene vicariance events (Riddle *et al.*, 2000; Jaeger *et al.*, 2005; Wilson & Pitts, 2010b). These geological events would have been particularly effective in dividing the ranges of species adapted to lower elevations. In contrast, isolation on mountain-top 'sky island' refugia during the Pleistocene appears to have shaped phylogeographical patterns in many alpine species (e.g. Knowles, 2000, 2001; DeChaine & Martin, 2005; Floyd *et al.*, 2005). Thus, an a priori hypothesis for low-elevation taxa is that divergence occurs on opposite sides of large geological barri-

Journal of Biogeography

ers, whereas one expectation for high-elevation taxa is that population divergence is associated with isolation on mountain tops. These hypotheses have been addressed with a number of species, but few studies have examined the phylogeography of taxa found across a range of elevations from the lowest valleys to the tops of the highest mountains.

Mountains are of particular interest with respect to population divergence, because organisms that occupy high elevations often exhibit adaptations to the harsh conditions and shorter spring and summer seasons associated with alpine environments (e.g. Fordyce & Nice, 2003; McCracken *et al.*, 2009; reviewed by Keller *et al.*, 2013). Selective pressures in the alpine zone can be strong enough to significantly reduce the up-slope migration of low-elevation individuals, restricting gene flow and leading to the genetic isolation of higher-elevation populations (e.g. Wilson *et al.*, 2013b). Organisms with broad elevational distributions can also form clines along elevational bands and diverge in parapatry (Endler, 1977). This pattern has been documented in a study of North American red foxes, which identified differentiation in a montane lineage isolated from low-elevation populations (Aubry *et al.*, 2009). In addition to clinal divergence, studies of organisms with wide elevational distributions have also revealed patterns of diversification attributed to orogeny-associated vicariance (e.g. Latch *et al.*, 2009). These studies illustrate the potentially complex and varying evolutionary and demographic histories that might be found in organisms with broad elevational ranges.

The sandhill skipper, *Polites sabuleti* (Boisduval, 1852), is a butterfly found throughout western North America. It displays a diversity of wing coloration and patterning across its range (Fig. 1). The distribution of geographical and ecological discontinuities in morphological variation has led taxonomists to separate *P. sabuleti* into at least 13 subspecies (Warren *et al.*, 2012) – including nine that occur in the state of Nevada (Austin, 1987) – making *P. sabuleti* one of the most morphologically diverse butterfly species in North America. Although subspecies designations may not correspond directly to distinct evolutionary lineages, they can be useful indicators of diversification (Phillimore *et al.*, 2007). Some subspecies of *P. sabuleti* also exhibit seasonal polyphenism in wing colour (e.g. *P. s. sabuleti*; Shapiro, 1975), adding to the breadth of wing pattern variation found within the species.

Polites sabuleti spans a wide elevational range, from sea level (Suisun Marsh in the San Francisco Bay area) to at least 3400 m (the White Mountains) in California. Most subspecies occur at low elevations (e.g. *P. s. sabuleti*), but some subspecies are alpine specialists (e.g. *P. s. tecumseh*). The extensive morphological diversity within *P. sabuleti* suggests recent or ongoing diversification and provides fertile ground for investigating restrictions in gene flow that might be associated with climatic gradients and geographical features.

In this study, we investigate patterns of population structure in *P. sabuleti* in order to evaluate different modes of mountain-associated differentiation. We predict that past geological events have had a strong influence on the diversification of *P. sabuleti*, as reflected in neutral genetic diversity and population structure. Specifically, we predict that low-elevation populations are genetically differentiated on opposite sides of major mountain ranges (resulting from montane vicariance), and high-elevation taxa have diverged on isolated mountain ranges (a 'sky island' model of divergence). Montane vicariance was observed in another species of skipper

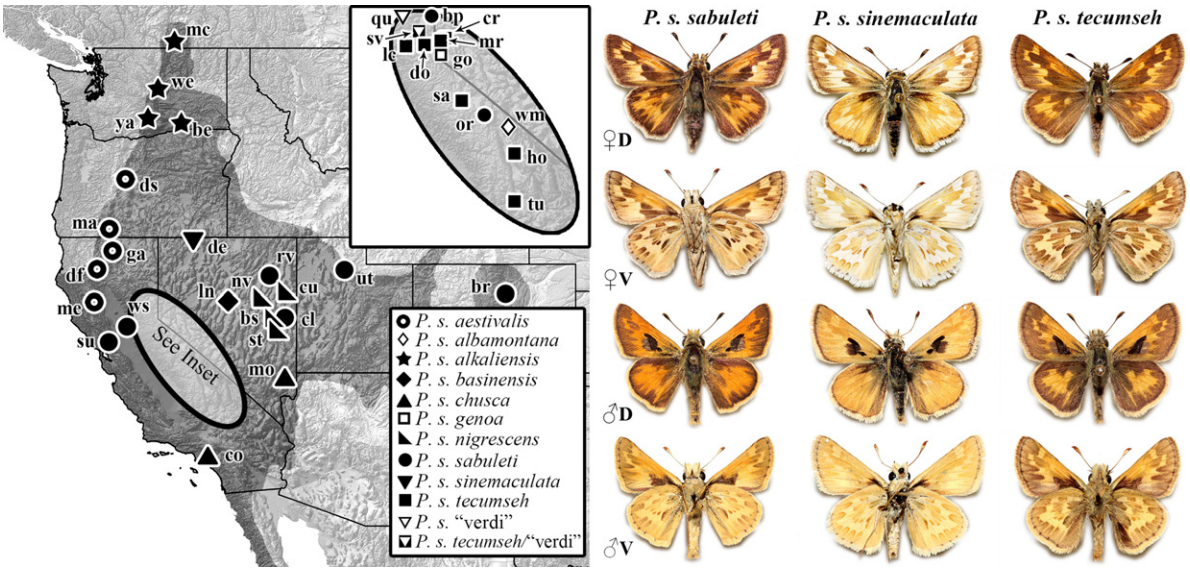

Figure 1 Subspecies designations for each sampled population of *Polites sabuleti*; the geographical range of *P. sabuleti* in western North America is shaded in dark grey (range roughly taken from Opler & Wright, 1999). To depict the degree of morphological variation among subspecies, dorsal (D) and ventral (V) wing surfaces of female and male skippers are displayed for three subspecies: *P. s. sabuleti*, *P. s. sinemaculata* and *P. s. tecumseh*. All photos are by Kim Davis, Mike Stangeland and Andrew Warren (Warren *et al.*, 2012).

with a similar distribution and ecology, but with a more limited elevational range (Forister *et al.*, 2004). Finally, we predict that the most morphologically and ecologically distinct populations, particularly those at high elevations, will show evidence of genetic differentiation with respect to adjacent populations in different habitats (a parapatric model of divergence). In this study, we use the terms divergence and differentiation to refer to an evolutionary process that leads to the origin of morphologically or genetically distinct populations or lineages (Mallet, 1995). Finally, we ask whether genetically distinct clusters of populations corroborate subspecies designations across the morphologically diverse set of *P. sabuleti* populations.

MATERIALS AND METHODS

Population sampling

Thirty-six populations of *P. sabuleti* were sampled across western North America, encompassing most of the species' geographical range (Fig. 2, Table 1, Table S1 in Appendix S1 of the Supporting Information) and elevational distribution. Attempts were made to collect 20 individuals from each population wherever possible. Eleven subspecies of *P. sabuleti* were represented in the sampled populations, including the nine subspecies Austin (1987) reported in the state of Nevada. *Polites sabuleti* 'verdi' is an informal designation reserved for ecologically and morphologically distinct individuals found at Quincy and Sierra Valley, California (A.M.S., pers. obs.). Unfortunately, the sole population of *P. s. albamontana* (White Mountains, California) failed to yield usable genetic data, so this subspecies was not included in our analyses. Specimens of *P. s. tecumseh* from Lang Crossing, CA, were collected in 1986 before the population was extirpated following the removal of grazing and subsequent changes in the vegetation (A.M.S., pers. obs.).

Mitochondrial DNA

DNA was extracted using the Purgene DNA Isolation Kit (Gentra Systems, Minneapolis, MN, USA). Once dried, the precipitated DNA was resuspended in 200 μL of H_20 and amplified using the primers EVA and PATRICK (C1-J-3038, C1-N-3782; Caterino & Sperling, 1999). A 611-bp region of the mitochondrial cytochrome *c* oxidase subunit II (*COII*) was sequenced from 189 individuals in 23 populations (Table 1). Electropherograms were visually inspected and sequences were aligned by hand in Sequencher 4.10.1 (Gene Codes, Ann Arbor, MI, USA).

To determine relationships among the mitochondrial haplotypes, a parsimony-based haplotype network was constructed from the *COII* sequences using tcs 1.21 (Clement *et al.*, 2000). Additionally, a Bayesian phylogenetic tree was constructed in beast 2.1.3 (Bouckaert *et al.*, 2014). For constructing the gene tree, a model of evolution was selected by comparing the Akaike information criterion (AIC) scores for

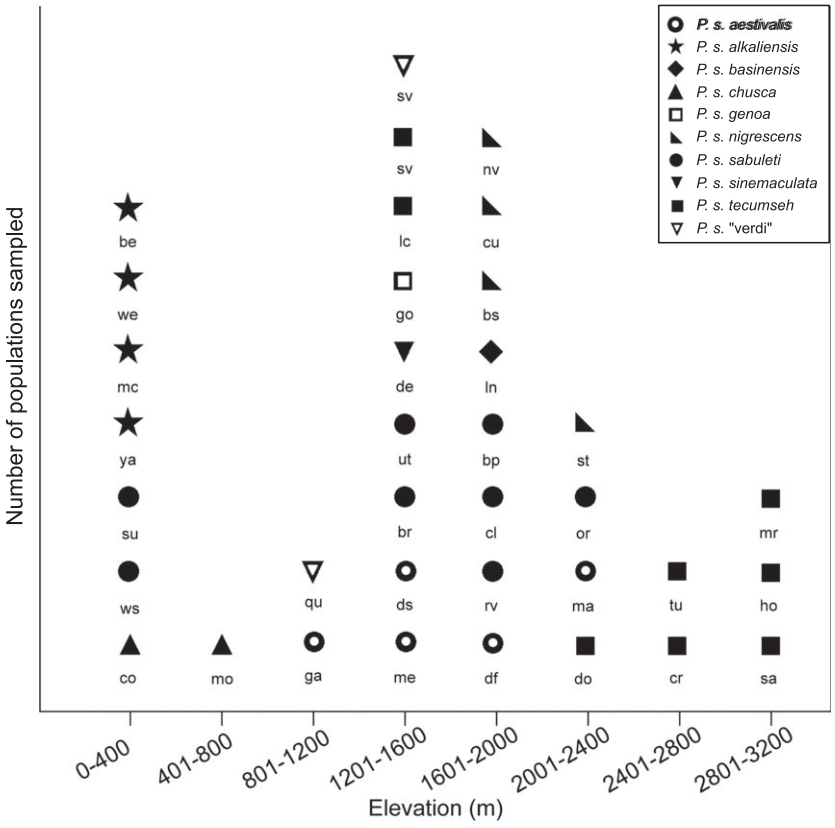

Figure 2 A histogram displaying the distribution both of sampling locations and of subspecies across the elevational range of *Polites sabuleti* in western North America. Two-letter locality codes correspond to those in Fig. 1 and Table 1.

Table 1 Sampling data from the 36 populations of *Polites sabuleti* from western North America included in this study, including population code, latitude (°), longitude (°) and elevation (m). Sample sizes (*n*) for each genetic protocol (either AFLP banding or mitochondrial *COII* sequencing) are given for each population. Three entities were found at Sierra Valley (*P. s. tecumseh*, *P. s.* 'verdi' and an unknown *P. s.* ssp.). For populations with AFLP sequences, the percentage of polymorphic loci with minor allele frequencies greater than 0.05 (%P) and expected heterozygosity (H_E) were calculated using AFLP-SURV 1.0 (Vekemans, 2002). All entities at Sierra Valley were pooled when calculating %P and H_E. Unbiased haplotype diversity, *h* (Nei, 1987), was calculated for populations with *COII* sequences.

					AFLP			*COII*	
Population	Code	Latitude, longitude	Elev. (m)	Subspecies	*n*	%P	H_E	*n*	*h*
Barr, CO	br	39.943, −104.759	1556	*sabuleti*	7	76.2	0.255	7	0.476
Bassett, NV	bs	39.434, −114.846	1907	*nigrescens*	8	71.4	0.259	0	n/a
Beckwourth Pass, CA	bp	39.793, −120.111	1605	*sabuleti*	15	88.1	0.278	0	n/a
Benton, WA	be	46.222, −119.145	107	*alkaliensis*	10	88.1	0.281	8	0.429
Carson, NV	cr	39.299, −119.914	2610	*tecumseh*	12	85.7	0.269	14	0.527
Cleveland, NV	cl	39.519, −114.514	1733	*sabuleti*	15	81.0	0.259	0	n/a
Corona, CA	co	33.875, −117.566	207	*chusca*	7	71.4	0.281	9	0.417
Currie, NV	cu	40.203, −114.696	1794	*nigrescens*	14	81.0	0.253	0	n/a
Deadfall, CA	df	41.236, −122.961	1841	*aestivalis*	14	76.2	0.277	0	n/a
Denio, NV	de	41.937, −118.696	1293	*sinemaculata*	16	54.8	0.201	11	0
Deschutes, OR	ds	43.703, −121.653	1310	*aestivalis*	8	73.8	0.281	10	0.2
Donner, CA	do	39.315, −120.349	2099	*tecumseh*	7	73.8	0.317	1	0
Gazelle, CA	ga	41.521, −122.519	845	*aestivalis*	0	n/a	n/a	7	0.524
Genoa, NV	go	38.997, −119.792	1429	*genoa*	5	71.4	0.264	0	n/a
Horseshoe, CA	ho	36.444, −118.169	3012	*tecumseh*	12	90.5	0.275	0	n/a
Lander, NV	ln	39.501, −117.119	1798	*basinensis*	8	76.2	0.263	4	0.833
Lang Crossing, CA	lc	39.312, −120.660	1470	*tecumseh*	5	69.0	0.272	4	0
Mendocino Pass, CA	me	39.794, −122.935	1527	*aestivalis*	18	81.0	0.281	11	0.327
Mica, BC	mc	49.014, −119.465	302	*alkaliensis*	5	100	0.282	3	0
Moapa, NV	mo	36.581, −114.469	406	*chusca*	0	n/a	n/a	8	0.429
Mount Ashland, OR	ma	42.079, −122.719	2194	*aestivalis*	21	71.0	0.236	0	n/a
Mount Rose, NV	mr	39.322, −119.930	2922	*tecumseh*	15	78.6	0.292	0	n/a
Newark Valley, NV	nv	39.597, −115.749	1788	*nigrescens*	15	66.7	0.241	0	n/a
Owens River, CA	or	37.688, −118.771	2078	*sabuleti*	8	76.2	0.261	0	n/a
Quincy, CA	qu	39.935, −120.929	1071	'verdi'	9	78.6	0.295	10	0.583
Ruby Valley, NV	rv	40.666, −115.140	1848	*sabuleti*	13	83.3	0.271	0	n/a
Saddlebag, CA	sa	37.965, −119.272	3090	*tecumseh*	0	n/a	n/a	9	0.417
Salt Lake, UT	ut	40.761, −111.890	1303	*sabuleti*	0	n/a	n/a	5	0.6
Sierra Valley, CA	sv	39.630, −120.361	1502	*tecumseh*	11	73.8	0.246	9	0.806
				'verdi'	15	73.8	0.246	11	1.036
				unknown ssp.	2	73.8	0.246	4	0.5
Steptoe, NV	st	38.851, −114.822	2204	*nigrescens*	0	n/a	n/a	3	0
Suisun Marsh, CA	su	38.231, −122.038	2	*sabuleti*	14	88.1	0.268	9	0.806
Tulare, CA	tu	35.850, −118.572	2509	*tecumseh*	0	n/a	n/a	4	0
Wenatchee, WA	we	47.424, −120.505	204	*alkaliensis*	14	88.1	0.292	8	0.536
West Sacramento, CA	ws	38.532, −121.569	10	*sabuleti*	15	78.6	0.259	10	0
White Mountains, CA	wm	37.464, −118.193	3422	*albamontana*	0	n/a	n/a	0	n/a
Yakima, WA	ya	46.602, −120.505	326	*alkaliensis*	9	73.8	0.250	10	0.389

models available in BEAST using jMODELTEST 2.1.5 (Darriba *et al.*, 2012). BEAST was run for one billion Markov chain Monte Carlo (MCMC) iterations, implementing a GTR+I model of evolution, a relaxed molecular clock rate (lognormal distribution; substitution rate of 1.15% per lineage per million years; Brower, 1994), and an initial burn-in of 10%. TRACER 1.6 was used to confirm chain convergence and TREEANNOTATOR 2.0.2 was used to find the maximum-clade-credibility tree.

Finally, a spatial analysis of molecular variance (SAMOVA) using the program SAMOVA 1.0 (Dupanloup *et al.*, 2002) was employed to document regional genetic structure between populations. SAMOVA uses a simulated annealing algorithm that maximizes the proportion of genetic variance (Φ_{CT}) partitioned among geographically contiguous populations to define a specified number of population clusters, which can then be used to infer phylogeographical patterns. SAMOVA was performed for each possible number of population clusters (*K*) from 2 to 20, with Φ_{CT} calculated for each value of *K*.

AFLP markers

Potentially problematic issues arise in molecular studies that only employ a single mitochondrial marker to infer

evolutionary histories, particularly in lineages that might be characterized by periods of isolation and hybridization (Forister *et al.*, 2008). We obtained AFLP profiles (Vos *et al.*, 1995) for 347 individuals from 29 populations of *P. sabuleti* (Table 1) as a complementary approach. One selective primer pair combination (MseI-CAGCA and EcoRI-ACA) was used to generate AFLP markers, and fluorescently labelled amplicons were separated and visualized on an ABI PRISM 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA) at the Nevada Genomics Center (Reno, NV). GeneMapper 4.0 was used to inspect, score and bin the AFLP markers following the methods of Gompert *et al.* (2010). To control for variation in amplification and sizing among the 96-well plates, one set of the same eight individuals was included as a subset of individuals on every plate. Any AFLP markers that appeared to exhibit inconsistent amplification across replicate runs or across individuals, or that otherwise could not be reliably scored, were conservatively excluded from analyses.

AFLP data were used to estimate the proportion of each individual's genome being assigned to each of the K population clusters using an admixture model in structure 2.3.2 (Pritchard *et al.*, 2000). When selecting parameters for structure, we followed the recommendations outlined by Gilbert *et al.* (2012) to increase the reproducibility of the results. Specifically, each MCMC algorithm used by the Bayesian model in structure was run for 1,000,000 iterations with an initial burn-in of 100,000 iterations, and 20 replicate MCMC trials were run for each level of K (from 1 to 11). The most probable number of population clusters was inferred by calculating the statistic delta-K (ΔK) with structure harvester 0.6.93 (Earl & vonHoldt, 2012), using the method described by Evanno *et al.* (2005). Assignment probabilities for individuals and populations were averaged across all 20 replicate MCMC trials for each value of K with clumpp 1.1.2 (Jakobsson & Rosenberg, 2007), which was run for 1000 random permutations using the Greedy algorithm. Average assignment probabilities were visualized with distruct 1.1 (Rosenberg, 2004). As a complementary clustering approach to structure, principal components analysis (PCA) was conducted on the correlation matrix of AFLP data using the 'princomp' function in R 2.14.1 (R Core Team, 2013).

An additional structure analysis was performed using the methods described above for 79 individuals from high-elevation Sierra Nevada populations (Carson, Donner, Horseshoe, Lang Crossing, Mount Rose and Sierra Valley) to investigate potential 'sky island' differentiation within the Sierra Nevada range. For this analysis, all subspecies found at Sierra Valley were included to determine whether multiple genetic clusters were present at this location where two subspecies co-occur (*P. s. tecumseh* and *P. s.* 'verdi'). Because there were only six high-elevation populations included in the second structure analysis, we analysed results from $K = 1$ to 7 (which allows values of ΔK to be calculated for values of K from 2 to 6).

Finally, a multiple regression on distance matrices (MRM; Lichstein, 2007) was conducted using the ecodist package (Goslee & Urban, 2007) in R to explicitly test for both isolation by distance and isolation by elevation among populations with AFLP data. The response variable in the MRM, pairwise F_{ST} between populations, was calculated from AFLP banding results using aflp-surv 1.0 (Vekemans, 2002). Pairwise geographical distance, net elevational difference and total elevation change between populations were measured in Google Earth (Google, Mountain View, CA; available at: http://www.google.com/earth/index.html) and used as predictor variables in the MRM (1000 permutations of the data were used to assess statistical significance at $\alpha = 0.05$). Net difference in elevation was calculated as the absolute value of the difference in elevation between two sites, whereas total elevational change was calculated as the sum of all elevational changes (both upwards and downwards) incurred when travelling in a straight line between two sites.

RESULTS

Twenty-five mitochondrial *COII* haplotypes were obtained from 189 individuals across 23 populations (Fig. 3) (GenBank accessions KJ147149–KJ147173). The haplotype network was divided into two major groups separated by five nucleotide substitutions (Fig. 4), with haplotypes P, K and R being the most common in one group and haplotypes C and B the most common in the other. Eighteen of the 25 haplotypes were private alleles found in only one population each (Fig. 3). The highest haplotype richness and diversity was found at Sierra Valley (Table 1), on the eastern side of the Sierra Nevada.

Relationships among haplotypes were consistent between the haplotype network and the phylogenetic tree. Based on a mitochondrial molecular clock of 1.15% substitutions per lineage per million years (Brower, 1994), *P. sabuleti* diverged from *Polites sonora c.* 2.82 Ma (95% credible interval, 1.32–4.50 Ma) (Fig. 4). The phylogenetic tree recovered strong support for a basal split from which haplotypes L, U and W diverged from the other haplotypes 814.5 ka (95% CI, 410.6–1273.7 ka). An additional clade composed of haplotypes associated with individuals mostly from low-elevation populations (E, G, F, C, D, X, A and B) diverged 656.2 ka (95% CI, 350.2–1093.3 ka), but received low support. The remaining two clades mostly consisted of haplotypes from individuals found in high-elevation Sierra Nevada populations (Q, O, Y, P and V) and northern California and Oregon (T, S, R, J, H, K and I), which diverged from each other 424.8 ka (95% CI, 178.0–709.3 ka). We note that coalescent time estimates for recently diverged populations with unknown demographic histories can be unreliable owing to the stochastic nature of mutation and the coalescent process; our estimates, and the uncertainty around these estimates, place the divergence of extant populations well within the Pleistocene.

In the SAMOVA analyses, Φ_{CT} appeared to reach a plateau in the amount of variation explained at $K = 15$ population clusters ($\Phi_{CT} = 0.780$; Fig. S1 in Appendix S2). Large increases in Φ_{CT} occurred at $K = 3$, 7 and 11 (see Fig. S2 in Appendix S2 for the population configurations inferred at

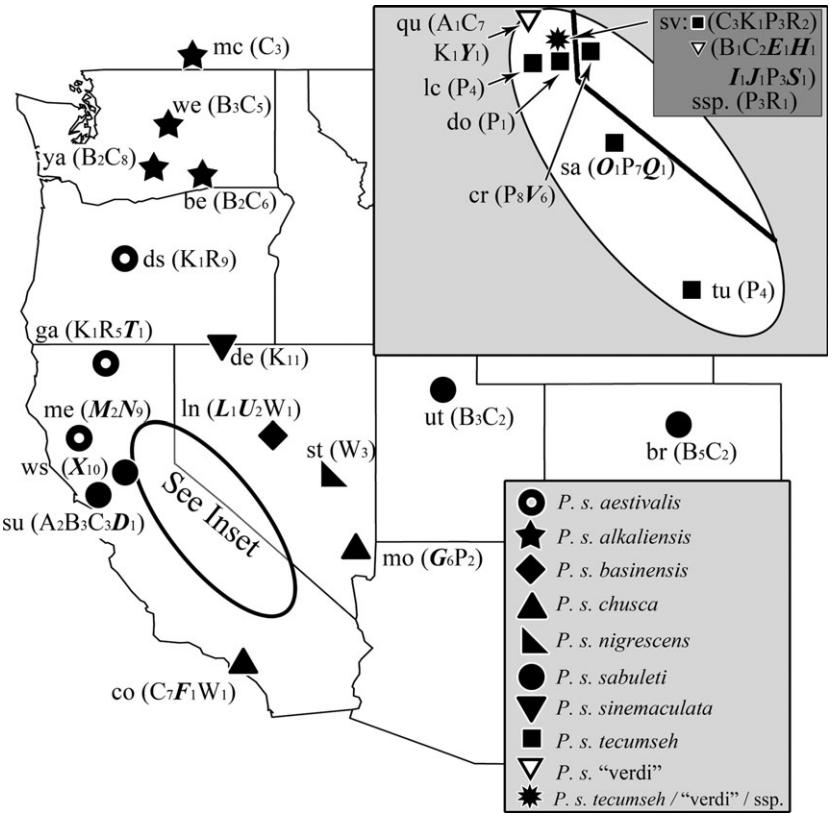

Figure 3 The distribution of *COII* haplotypes among 23 populations of *Polites sabuleti* in western North America. Each capital letter corresponds to a unique haplotype sequence and each subsequent number represents the number of individuals with that haplotype in the population. Haplotypes at Sierra Valley (sv) are grouped by nominal taxonomy because three entities are found at that population (*P. s. tecumseh*, *P. s.* 'verdi', and an unknown *P. s.* ssp.). Private alleles (haplotypes found at only one location) are bold and italicized.

each of these levels). For simplicity, Fig. 5 shows only the configuration for $K = 7$, because higher levels of K become less informative as additional populations are isolated (i.e. in clusters with a single population). The most widespread population cluster includes eight low-elevation populations from California, Colorado, Utah and Washington and encompasses the subspecies *P. s. alkaliensis*, *P. s. chusca* and *P. s. sabuleti* (Fig. 5). A high-elevation cluster in the Sierra Nevada of California and Nevada was also delineated by the SAMOVA and encompassed all populations of *P. s. tecumseh* and one population of *P. s. aestivalis* (Mendocino Pass in the North Coast Range of California, Fig. 5). The remaining two populations of *P. s. aestivalis* were grouped in a separate cluster with the sole population of *P. s. sinemaculata*. In addition, two populations from central Nevada were clustered together in the SAMOVA, even though they contained separate subspecies (*P. s. basinensis* and *P. s. nigrescens*). The Quincy (*P. s.* 'verdi'), Moapa (*P. s. chusca*) and West Sacramento (*P. s. sabuleti*) populations were each isolated in single-population clusters. Table 2 contains AMOVA results associated with the SAMOVA at $K = 7$. Most of the variation in molecular variance was explained among the seven regions (67.75%), whereas 22.63% of the variation was explained within populations and the remaining 9.63% of the variation was explained among populations within regions.

For the AFLP profiling, 42 anonymous markers were generated for 347 individuals from 29 populations. The largest peak of ΔK in the STRUCTURE analysis was at $K = 5$ (Fig. S3 in Appendix S2), suggesting the existence of five genetic clusters in *P. sabuleti* (Fig. 6). Although many populations include individuals assigned to different clusters (at both $K = 5$ and $K = 2$, discussed below), we found both geographical and taxonomic structure among populations. Of the 29 populations included in the STRUCTURE analysis, six populations were clustered together to form a high-elevation Sierra Nevada cluster (coloured yellow in Fig. 6) that included all of the populations of *P. s. tecumseh*. In addition, five populations from northern California and Oregon were clustered together (red in Fig. 6), encompassing all populations of *P. s. aestivalis* and the sole population of *P. s. sinemaculata*. A number of low-elevation populations also were clustered together (purple in Fig. 6), corresponding to the widespread low-elevation cluster found in the SAMOVA results (see grouping in Fig. 5 denoted by a solid line). The remaining populations were predominantly composed of two relatively heterogeneous clusters of mostly lower-elevation subspecies (blue and green in Fig. 6). Individuals with high assignment probabilities to either the blue or the green genetic cluster were scattered across many of the other populations. A smaller ΔK peak was found in the STRUCTURE results at $K = 2$ (Fig. S3). The two genetic clusters implied by this peak complement the results from the $K = 5$ analysis, with one cluster matching the red and yellow clusters, and the other corresponding to the blue, green and purple clusters (Fig. 6). Results from STRUCTURE were generally concordant with the PCA analysis (Fig. S4 in Appendix S2), which

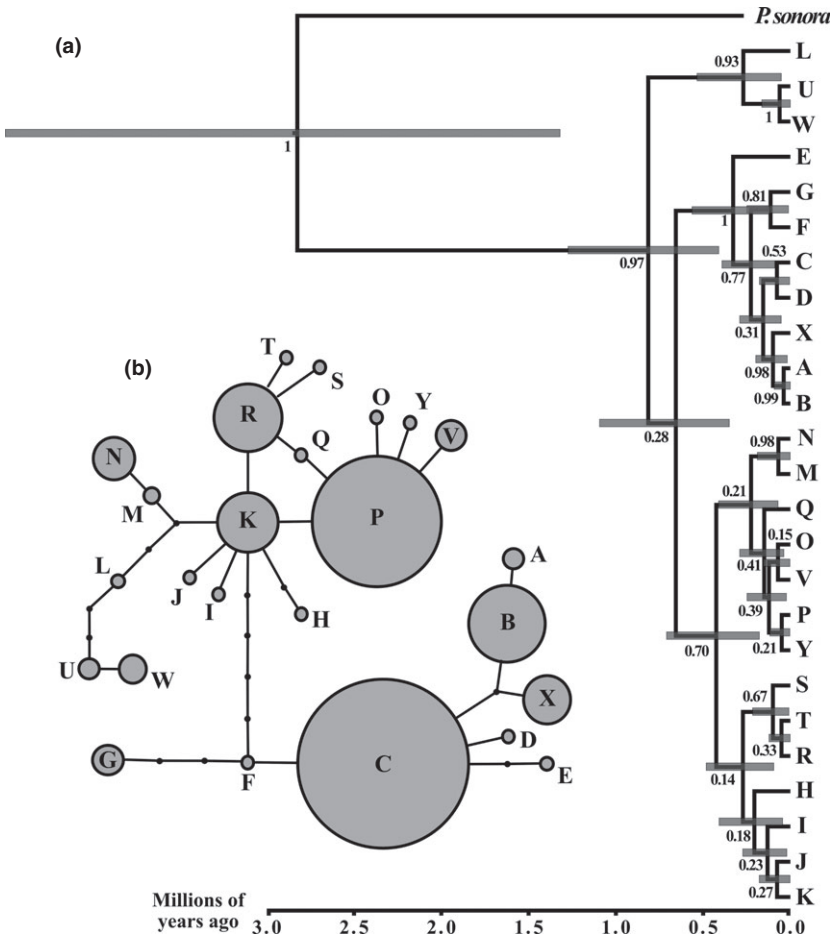

Figure 4 (a) A phylogenetic tree displays the relationships among *Polites sabuleti* haplotypes and diversification patterns in western North America across time. Node labels correspond to posterior probabilities and node bars show the 95% credible interval for the date of divergence. (b) The genetic relationships among *COII* haplotype sequences are shown in a parsimony-based haplotype network. The distance between each circle represents one nucleotide substitution, with small, unlabelled circles depicting unsampled intermediate sequences. For scale, the sizes of each labelled circle are proportional to the number of individuals with that haplotype; the smallest labelled circles represent a sample size of one and the largest circle (C) represents 48 individuals.

separated *P. s. tecumseh* and *P. s. aestivalis* individuals from a large cluster containing individuals from the other subspecies; *P. s.* 'verdi' individuals were intermediate between *P. s. tecumseh* and the main cluster.

In the STRUCTURE analysis of the six high-elevation *P. s. tecumseh* populations, ΔK peaked at $K = 2$ (Fig. S5 in Appendix S2), and all of the individuals had admixture proportions roughly split between the two genetic clusters (Fig. S6 in Appendix S2), suggesting an absence of high-elevation population structure. Finally, for the MRM that tested for isolation by distance and by elevation, none of the distance measures (pairwise geographical distance, net elevational difference and total elevational change) were significant predictors of pairwise F_{ST} values calculated from AFLPs ($R^2 = 0.02$; $P = 0.69$).

DISCUSSION

Regional population structure

Substantial population structure was found throughout the range of *P. sabuleti* across western North America (Figs 4 & 6). The most basal intraspecific split in the phylogenetic tree (Fig. 4) was dated to more than 800,000 years ago, suggesting that differentiation in *P. sabuleti* occurred considerably after the rise of the Sierra Nevada mountains 2–15 Ma (Wilson & Pitts, 2010a) and could have been affected by Pleistocene glacial cycles. In the SAMOVA, mitochondrial data provided evidence for four main genetic clusters: a high-elevation Sierra Nevada cluster, a northern California and Oregon cluster, a widespread low-elevation cluster, and a central Nevada cluster.

The low-elevation population cluster identified by STRUCTURE analysis of AFLP data (blue in the $K = 2$ plot in Fig. 6) largely corresponds to the most geographically widespread cluster and the Nevada cluster identified by SAMOVA for mitochondrial DNA (Fig. 5). Because many of the low-elevation populations contained individuals assigned to weakly delimited low-elevation genetic clusters (blue and green in the $K = 5$ plot in Fig. 6), the evolutionary histories of these populations remain cloudy, perhaps suggesting some history of geographical separation followed by gene flow. Substantial gene flow is indicated in particular by populations with individuals that are wholly or mostly assigned to different clusters. For example, we infer that a location such as Mount Ashland, Oregon, has received historical or recent gene flow from other locations (Fig. 6; this is evident at both $K = 2$ and $K = 5$). Care must be taken when interpreting results from STRUCTURE analyses where $K = 2$, because genetic artefacts can lead to the detection of false population

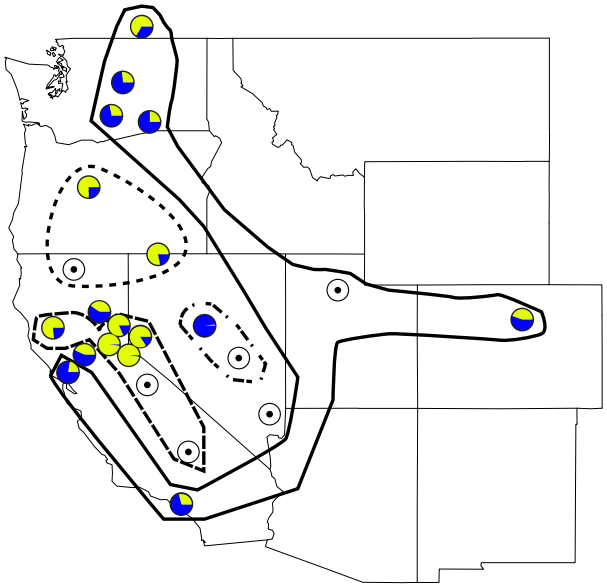

Figure 5 Detailed results of spatial analysis of molecular variance (SAMOVA) for seven population clusters of *Polites sabuleti* in western North America. Each population cluster is depicted with a distinct line type. Clusters with only one population are not circled. For populations that were also included in the STRUCTURE analysis (see Fig. 5), pie charts display the population level assignment values from the $K = 2$ analysis (yellow corresponds to a high-elevation genetic cluster and blue corresponds to a low-elevation cluster). Populations that were not analysed in STRUCTURE are denoted with a black dot in a white circle.

structure, even when individuals are genetically panmictic (i.e. when the true answer is $K = 1$; Pritchard *et al.*, 2010). Despite this, it is reasonable to infer that the population structure at $K = 2$ is biologically meaningful because the results closely match the patterns and interpretations from $K = 5$.

In the analysis of molecular variance (AMOVA) (extracted from the $K = 7$ SAMOVA output), most of the molecular variation in mitochondrial DNA was explained by variation among regions (Table 2), which is consistent with the existence of regional differentiation. The high proportion of regional structure found in *P. sabuleti* mitochondrial DNA is similar to results found in *Hesperia comma*, another western North American skipper species complex (Forister *et al.*, 2004). In contrast, far less genetic diversity and regional structure were found in the *Pyrgus communis/albescens* skipper species complex across a wider geographical range: a single haplotype in that complex was common from the west coast of North America to the east (Fordyce *et al.*, 2008).

Sample size should always be considered when interpreting population genetic results. Given the number of specimens that can be processed, there are trade-offs between sampling many populations and sampling intensely within populations. In this study, we chose to maximize geographical coverage. With more intense within-population sampling, rare alleles may have been revealed that could have changed our inferences of population structure. However, rare variants are typically less informative when analysing population structure and are better suited for examining more recent evolutionary processes (Gompert *et al.*, 2014). Thus, greater resolution of population-level differences and similarities would be unlikely to overturn the inferences that we have drawn.

Evidence for subspecies delimitations

The subspecies has been defined as a group of morphologically, geographically and genetically distinct populations that differ from other such groups within the same species (Mayr, 1942). Subspecies designations do not always correspond to evolutionary history, however, with some studies finding a lack of genetic evidence for subspecific hypotheses (e.g. Ball & Avise, 1992; Zink, 2004; Wilson & Pitts, 2008). Despite these inconsistencies, subspecies designations can be useful as hypotheses of divergence in evolutionary biology, and it should be noted that the role of subspecies in conservation planning and prioritization of management actions is highly contentious (Phillimore & Owens, 2006; Braby *et al.*, 2012).

In this study, we find genetic evidence supporting some, but not all, subspecies of *P. sabuleti*. For example, populations of the subspecies *P. s. tecumseh* may constitute a distinct genetic cluster, as suggested by nuclear DNA and to a lesser extent mitochondrial DNA (Figs 4 & 5). In the $K = 5$ STRUCTURE plot (Fig. 6), individuals of *P. s. tecumseh* were predominately assigned to the yellow genetic cluster, although some populations did include individuals assigned to more than one cluster. In the SAMOVA results, the populations of *P. s. tecumseh* also cluster together (the Sierra Nevada cluster in Fig. 5), but that cluster also includes one *P. s. aestivalis* population (Mendocino Pass). This result is interesting in the light of morphological observations that *P. s. aestivalis* is phenotypically similar to *P. s. tecumseh* (Emmel *et al.*, 1998), and suggests that there might be a correlation between morphological and molecular differentiation across some subspecies of *P. sabuleti*.

Table 2 Analysis of molecular variance (AMOVA) results for mitochondrial *COII* sequences from 36 populations of *Polites sabuleti* from western North America. Regions and populations match those found in Fig. 5.

Source of variation	d.f.	Sum of squares	Variance	% variation	*P*	Φ-statistics
Among regions	6	321.766	2.136	67.75	< 0.001	$\Phi_{CT} = 0.678$
Among populations within regions	16	49.274	0.304	9.63	< 0.001	$\Phi_{SC} = 0.298$
Within populations	166	118.431	0.713	22.63	< 0.001	$\Phi_{ST} = 0.774$

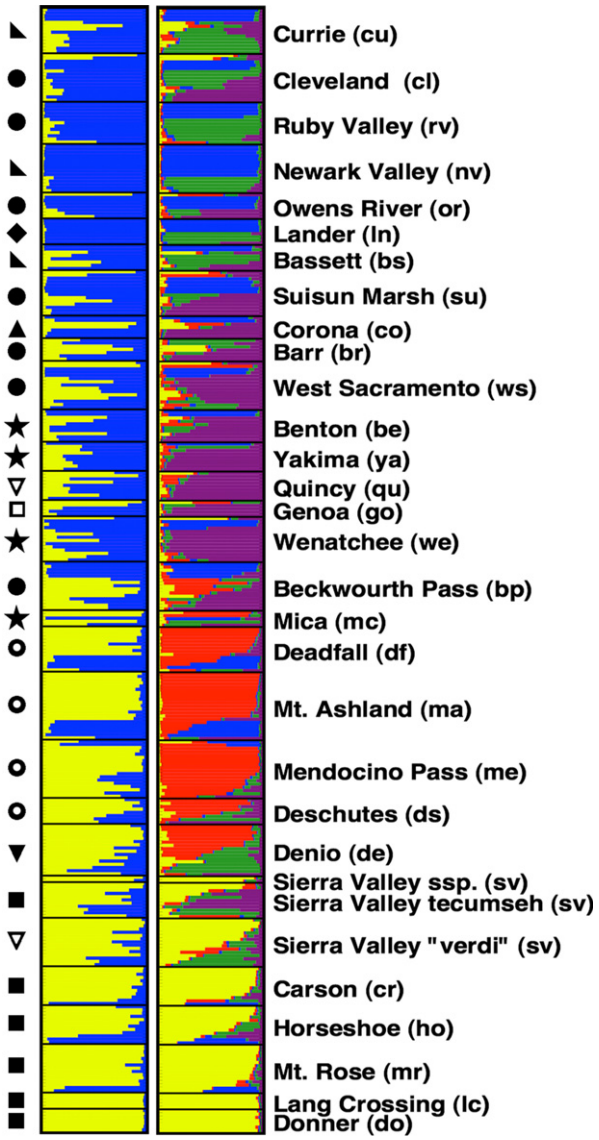

Figure 6 STRUCTURE plots depicting the results from analyses including $K = 2$ and $K = 5$ population clusters for *Polites sabuleti* in western North America. Each bar represents one individual and each colour represents one of the five population clusters. The proportion of a colour in each bar is proportional to the individual's assignment probability to the corresponding population cluster. Subspecies symbols are the same as those used in Figs 1 & 2.

In contrast, we found no evidence suggesting that *P. s. sinemaculata* is a distinct entity despite the unique morphology of these individuals (Fig. 1). *Polites s. sinemaculata* was recently considered for protection under the Endangered Species Act and found to be not warranted for listing by the U.S. Fish and Wildlife Service (2012) owing to a lack of threats to its persistence (e.g. habitat destruction or overutilization). In the SAMOVA analysis, the sole population of *P. s. sinemaculata*, Denio, grouped with two populations of *P. s. aestivalis* in a northern California/Oregon cluster (Fig. 5). Additionally, individuals of *P. s. sinemaculata* were assigned to the yellow cluster in the $K = 2$ STRUCTURE analysis, (as were most individuals of *P. s. tecumseh* and *P. s. aestivalis*; Fig. 6), whereas in the $K = 5$ analysis, *P. s. sinemaculata* individuals showed large assignment probabilities to either the red cluster (to which many *P. s. aestivalis* individuals were assigned) or to the green cluster (to which some low-elevation individuals were assigned).

Subspecific designations for other low-elevation subspecies were also ambiguous from the perspective of population-genetic evidence, as populations from multiple subspecies generally clustered together in SAMOVA and were heterogeneous mixtures of genetic clusters in STRUCTURE. The populations from central and eastern Nevada (designated as *P. s. basinensis* and *P. s. nigrescens*) might compose at least one distinct gene pool as suggested by the SAMOVA results (Fig. 5), but more geographically and genetically intensive sampling is needed to confirm any taxonomic inferences and to parse out relationships among the subspecies, in particular *P. s. alkaliensis*, *P. s. chusca* and *P. s. sabuleti*. In general, results from *P. sabuleti* support a perspective that subspecies can in some cases be useful as indications of genetic divergence and thus can be useful in biogeography, but such divergence should not be assumed without extensive multilocus investigation.

Phylogeographical patterns

In many low-elevation North American plants and animals, regional structure is evident on opposite sides of large mountain ranges, such as the Rocky Mountains, Sierra Nevada and Transverse Ranges (e.g. Calsbeek *et al.*, 2003; Forister *et al.*, 2004; Feldman & Spicer, 2006; Chatzimanolis & Caterino, 2007; Oliver & Shapiro, 2007). Instead of a simple pattern of genetic divergence on opposite sides of mountains, we discovered a large proportion of private mitochondrial haplotypes throughout the range of *P. sabuleti* (Fig. 3), a pattern that was also documented across the range of the pallid-dotted blue butterfly (*Euphilotes pallescens*) in western North America (Wilson *et al.*, 2013a). This result is consistent with the importance of mountains as barriers, but more generally suggests that gene flow might be localized in *P. sabuleti*, regardless of geological features.

Interestingly, *P. sabuleti* is considered a strong flier and can feed on a variety of common grasses (Opler & Wright, 1999; Shapiro & Manolis, 2007), so the possibility of locally restricted gene flow among populations is somewhat surprising, although perhaps foreshadowed by the great morphological diversity observed throughout the species' range (Fig. 1). In addition, it is quite possible that *P. sabuleti* has a more specialized diet or habitat preference at the population level (Fox & Morrow, 1981), which could explain the high levels of localized gene flow in low-elevation populations. For instance, the Sierra Valley location at the northern range limit of *P. s. tecumseh* had the highest mitochondrial haplotype diversity found in the study (10 of the 25 haplotypes were present at this single location; Table 1; Fig. 3). The

high haplotype diversity found at Sierra Valley contrasts strongly with a phylogeographical study of Sierra Nevada butterflies in the *Oeneis chryxus* complex that found a latitudinal trend in haplotype diversity (with southern populations being more diverse than northern populations), which was perhaps the result of a large Pleistocene refugium found in the southern Sierra Nevada (Nice & Shapiro, 2001). Sierra Valley is also noteworthy in having two morphologically distinguishable entities, *P. s. tecumseh* and *P. s.* 'verdi', co-occurring in sympatry at a fine geographical scale. Despite our expectations based on morphological differences for both Sierra Valley entities, the mitochondrial haplotypes were not partitioned by phenotype (Fig. 4) and the results from the STRUCTURE analysis on AFLP data yielded qualitatively similar results (Fig. 6). Although both mitochondrial and nuclear DNA evidence contradict the genetic distinctness of *P. s.* 'verdi', more powerful molecular techniques may shed additional light on the apparent coexistence of morphologically distinct individuals at Sierra Valley.

In contrast to vicariance associated with mountain ranges, population isolation on mountain tops can induce divergence in alpine organisms, resulting in sky-island patterns of diversification (e.g. Knowles, 2000, 2001; DeChaine & Martin, 2005). In *P. sabuleti*, our results indicate that the high-elevation Sierra Nevada populations constitute a single genetic lineage separated to some extent from low-elevation lineages (Figs 4 & 6). In Fig. 5, for example, the high-elevation populations identified in the STRUCTURE analysis (yellow in the $K = 5$ plot; Fig. 6) correspond to a mostly Sierra Nevada cluster identified in the SAMOVA analysis. This is consistent with the results of Shapiro (1975), who found evidence that the morphological differentiation between low-elevation *P. s. sabuleti* and high-elevation *P. s. tecumseh* was genetically based and that isolation was maintained by differences in phenology. Contrary to sky-island models of divergence found in other alpine insect lineages in the Sierra Nevada (e.g. Schoville & Roderick, 2010; Schoville *et al.*, 2012), little structure was found among high-elevation populations of *P. s. tecumseh* (Fig. S6), suggesting a lack of dispersal limitation among alpine locations (or recent expansion across alpine locations from a single source). These results are also consistent with a single colonization of the alpine zone by *P. sabuleti*. The lack of genetic structure among *P. s. tecumseh* populations presents an interesting dichotomy from the appearance of restricted gene flow among low-elevation populations and suggests that vicariance among isolated mountain tops has not been a dominant driver of diversification in this species.

Elevational differentiation in *P. sabuleti*

Although we did not find support for genetic isolation on different mountain tops, we did find evidence that some of the most morphologically and ecologically distinct *P. sabuleti* populations at high elevations are isolated from lower-elevation populations. Specifically, the non-significant effect of elevation on F_{ST} (MRM; see Results) suggests that the relationship between elevation and genetic divergence is not continuous across populations of *P. sabuleti*. Instead, the effect of elevation appears to be associated only with populations at the highest elevations, which is qualitatively apparent (e.g. the clustering of *P. s. tecumseh* populations in Fig. 6) even if not detected in MRM results. We cannot yet distinguish among biological or biogeographical mechanisms associated with the putative isolation of the high-elevation populations, however (e.g. Lozier *et al.*, 2013). In particular, it is difficult to address the possibility that adaptation to the high-elevation environment acts to restrict gene flow. Furthermore, it can be difficult to distinguish isolation by distance from restricted gene flow associated with ecological selection against immigrants (Frantz *et al.*, 2009). It is interesting to note that Keller *et al.* (2013) only found a significant relationship between genetic distance and elevation across taxa when examining adaptively relevant coding loci and not neutral loci, so the non-significant result of the MRM in this study may be due to limitations of the genetic markers used (AFLPs). A future study incorporating additional genetic markers in a model-based framework (e.g. Nice *et al.*, 2013) will therefore be needed to confidently ascertain the role of ongoing gene flow for the differentiation for *P. s. tecumseh*.

Isolation between *P. s. tecumseh* and low-elevation populations might be reduced in the future, as many Californian butterflies have experienced upward shifts in their distributional ranges in response to climate change (Forister *et al.*, 2010), possibly resulting in an increased likelihood of gene flow between low- and high-elevation populations. Shifting regional climate patterns could also affect gene flow between parapatric or sympatric entities, as has been observed in hybridizing *Colias* butterflies (Jahner *et al.*, 2012). In fact, although many of the highest populations of *P. s. tecumseh* were almost entirely composed of the high-elevation gene pool (e.g. Mount Rose and Donner Pass), lower-elevation sites already show evidence of gene flow from low-elevation genetic clusters (e.g. Sierra Valley; Fig. 6).

In sum, the molecular patterns found here in *P. sabuleti* suggest a complicated history of differentiation associated with mountains in western North America. Although we found little support for vicariant divergence among genetic clusters on opposite sides of mountain ranges or on the tops of isolated ranges, we did find support for the differentiation of high-elevation Sierra Nevada populations from lower-elevation populations. Thus, the phylogeographical patterns in *P. sabuleti* both highlight the complexity of biogeographical histories in the region and support the role of elevational gradients as drivers of diversification.

ACKNOWLEDGEMENTS

We thank G. Austin, B. Gendron, Z. Gompert, M. Hofmann, E. Leger, M. Robinson, E. Runquist, J. Scott and D. Threatful for providing specimens, Z. Gompert for analysing AFLP data,

and A. Warren for use of *P. sabuleti* photos. Six anonymous referees, Z. Gompert, A. Hornsby, C. Feldman, G. Hoelzer and the University of Nevada, Reno, EECB peer review group provided invaluable insight. J.P.J. was supported by National Science Foundation DEB-1145609; M.L.F. was supported by the University of Nevada, Reno Biology Department; C.C.N. was supported by a Texas State Research Enhancement Grant; J.A.F. and Z.H.M. were supported by the Department of Ecology and Evolutionary Biology at the University of Tennessee, and A.M.S. was supported by California Agricultural Experiment Station Project CA-D*-AZO-3994-H, 'Climatic Range Limitation of Phytophagous Lepidopterans'.

REFERENCES

Aubry, K.B., Statham, M.J., Sacks, B.N., Perrine, J.D. & Wisely, S.M. (2009) Phylogeography of the North American red fox: vicariance in Pleistocene forest refugia. *Molecular Ecology*, **18**, 2668–2686.

Austin, G.T. (1987) Nevada populations of *Polites sabuleti* and the descriptions of five new subspecies. *Bulletin of the Allyn Museum*, **109**, 1–24.

Avise, J.C. (2004) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA.

Ball, R.M. & Avise, J.C. (1992) Mitochondrial DNA phylogeographic differentiation among avian populations and the evolutionary significance of subspecies. *The Auk*, **109**, 626–636.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond, A.J. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Braby, M.F., Eastwood, R. & Murray, N. (2012) The subspecies concept in butterflies: has its application in taxonomy and conservation biology outlived its usefulness? *Biological Journal of the Linnean Society*, **106**, 699–716.

Brower, A.V.Z. (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences USA*, **91**, 6491–6495.

Calsbeek, R., Thompson, J.N. & Richardson, J.E. (2003) Patterns of molecular evolution and diversification in a biodiversity hotspot: the California Floristic Province. *Molecular Ecology*, **12**, 1021–1029.

Caterino, M.S. & Sperling, F.A.H. (1999) *Papilio* phylogeny based on mitochondrial cytochrome oxidase I and II genes. *Molecular Phylogenetics and Evolution*, **11**, 122–137.

Chatzimanolis, S. & Caterino, M.S. (2007) Toward a better understanding of the "Transverse Range break": lineage diversification in southern California. *Evolution*, **61**, 2127–2141.

Clement, M., Posada, D. & Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

DeChaine, E.G. & Martin, A.P. (2005) Historical biogeography of two alpine butterflies in the Rocky Mountains: broad-scale concordance and local-scale discordance. *Journal of Biogeography*, **32**, 1943–1956.

Dupanloup, I., Schneider, S. & Excoffier, L. (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**, 2571–2581.

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**, 359–361.

Emmel, J.F., Emmel, T.C. & Mattoon, S.O. (1998) New subspecies of Hesperiidae (Lepidoptera) from California. *Systematics of western North American butterflies* (ed. by T.C. Emmel), pp. 201–206. Mariposa Press, Gainesville, FL.

Endler, J.A. (1977) *Geographic variation, speciation, and clines*. Princeton University Press, Princeton, NJ.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Feldman, C.R. & Spicer, G.S. (2006) Comparative phylogeography of woodland reptiles in California: repeated patterns of cladogenesis and population expansion. *Molecular Ecology*, **15**, 2201–2222.

Floyd, C.H., Van Vuren, D.H. & May, B. (2005) Marmots on Great Basin mountaintops: using genetics to test a biogeographic paradigm. *Ecology*, **86**, 2145–2153.

Fordyce, J.A. & Nice, C.C. (2003) Variation in butterfly egg adhesion: adaptation to local host plant senescence characteristics? *Ecology Letters*, **6**, 23–27.

Fordyce, J.A., Forister, M.L., Nice, C.C., Burns, J.M. & Shapiro, A.M. (2008) Patterns of genetic variation between the checkered skippers *Pyrgus communis* and *Pyrgus albescens* (Lepidoptera: Hesperiidae). *Annals of the Entomological Society of America*, **101**, 794–800.

Forister, M.L., Fordyce, J.A. & Shapiro, A.M. (2004) Geological barriers and restricted gene flow in the holarctic skipper *Hesperia comma* (Hesperiidae). *Molecular Ecology*, **13**, 3489–3499.

Forister, M.L., Nice, C.C., Fordyce, J.A., Gompert, Z. & Shapiro, A.M. (2008) Considering evolutionary processes in the use of single-locus genetic data for conservation, with examples from the Lepidoptera. *Journal of Insect Conservation*, **12**, 37–51.

Forister, M.L., McCall, A.C., Sanders, N.J., Fordyce, J.A., Thorne, J.H., O'Brien, J., Waetjen, D.P. & Shapiro, A.M. (2010) Compounded effects of climate change and habitat alteration shift patterns of butterfly diversity. *Proceedings of the National Academy of Sciences USA*, **107**, 2088–2092.

Fox, L.R. & Morrow, P.A. (1981) Specialization: species property or local phenomenon? *Science*, **211**, 887–893.

Frantz, A.C., Cellina, S., Krier, A., Schley, L. & Burke, T. (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, **46**, 493–505.

Gilbert, K.J., Andrew, R.L., Bock, D.G., Franklin, M.T., Kane, N.C., Moore, J.-S., Moyers, B.T., Renaut, S., Rennison, D.J., Veen, T. & Vines, T.H. (2012) Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program STRUCTURE. *Molecular Ecology*, **21**, 4925–4930.

Gompert, Z., Lucas, L.K., Fordyce, J.A., Forister, M.L. & Nice, C.C. (2010) Secondary contact between *Lycaeides idas* and *L. melissa* in the Rocky Mountains: extensive admixture and a patchy hybrid zone. *Molecular Ecology*, **19**, 3171–3192.

Gompert, Z., Lucas, L.K., Buerkle, C.A., Forister, M.L., Fordyce, J.A. & Nice, C.C. (2014) Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, **23**, 4555–4573.

Goslee, S.C. & Urban, D.L. (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**, 1–19.

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hewitt, G.M. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 183–195.

Jaeger, J.R., Riddle, B.R. & Bradford, D.F. (2005) Cryptic Neogene vicariance and Quaternary dispersal of the red-spotted toad (*Bufo punctatus*): insights on the evolution of North American warm desert biotas. *Molecular Ecology*, **14**, 3033–3048.

Jahner, J.P., Shapiro, A.M. & Forister, M.L. (2012) Drivers of hybridization in a 66-generation record of *Colias* butterflies. *Evolution*, **66**, 818–830.

Jakobsson, M. & Rosenberg, N.A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Keller, I., Alexander, J.M., Holderegger, R. & Edwards, P.J. (2013) Widespread phenotypic and genetic divergence along altitudinal gradients in animals. *Journal of Evolutionary Biology*, **26**, 2527–2543.

Knowles, L.L. (2000) Tests of Pleistocene speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of western North America. *Evolution*, **54**, 1337–1348.

Knowles, L.L. (2001) Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Molecular Ecology*, **10**, 691–701.

Latch, E.K., Heffelfinger, J.R., Fike, J.A. & Rhodes, O.E. (2009) Species-wide phylogeography of North American mule deer (*Odocoileus hemionus*): cryptic glacial refugia and postglacial recolonization. *Molecular Ecology*, **18**, 1730–1745.

Lichstein, J.W. (2007) Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecology*, **188**, 117–131.

Lozier, J.D., Strange, J.P. & Koch, J.B. (2013) Landscape heterogeneity predicts gene flow in a widespread polymorphic bumble bee, *Bombus bifarius* (Hymenoptera: Apidae). *Conservation Genetics*, **14**, 1099–1110.

Mallet, J. (1995) A species definition for the modern synthesis. *Trends in Ecology and Evolution*, **10**, 294–299.

Mayr, E. (1942) *Systematics and the origin of species*. Columbia University Press, New York.

McCracken, K.G., Barger, C.P., Bulgarella, M., Johnson, K.P., Sonsthagen, S.A., Trucco, J., Valqui, T.H., Wilson, R.E., Winker, K. & Sorenson, M.D. (2009) Parallel evolution in the major haemoglobin genes of eight species of Andean waterfowl. *Molecular Ecology*, **18**, 3992–4005.

Nei, M. (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.

Nice, C.C. & Shapiro, A.M. (2001) Patterns of morphological, biochemical, and molecular evolution in the *Oeneis chryxus* complex (Lepidoptera: Satyridae): a test of historical biogeographical hypotheses. *Molecular Phylogenetics and Evolution*, **20**, 111–123.

Nice, C.C., Gompert, Z., Fordyce, J.A., Forister, M.L., Lucas, L.K. & Buerkle, C.A. (2013) Hybrid speciation and independent evolution in lineages of alpine butterflies. *Evolution*, **67**, 1055–1068.

Oliver, J.C. & Shapiro, A.M. (2007) Genetic isolation and cryptic variation within the *Lycaena xanthoides* species group (Lepidoptera: Lycaenidae). *Molecular Ecology*, **16**, 4308–4320.

Opler, P.A. & Wright, A.B. (1999) *Western butterflies*. Houghton Mifflin, Boston, MA.

Phillimore, A.B. & Owens, I.P.F. (2006) Are subspecies useful in evolutionary and conservation biology? *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1049–1053.

Phillimore, A.B., Orme, C.D.L., Davies, R.G., Hadfield, J.D., Reed, W.J., Gaston, K.J., Freckleton, R.P. & Owens, I.P.F. (2007) Biogeographical basis of recent phenotypic divergence among birds: a global study of subspecies richness. *Evolution*, **61**, 942–957.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Pritchard, J.K., Wen, X. & Falush, D. (2010) *Documentation for* structure *software: version 2.3*. Available at http://pritch.bsd.uchicago.edu/structure.html.

R Core Team (2013) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.r-project.org/.

Riddle, B.R., Hafner, D.J. & Alexander, L.F. (2000) Phylogeography and systematics of the *Peromyscus eremicus* species group and the historical biogeography of North American warm regional deserts. *Molecular Phylogenetics and Evolution*, **17**, 145–160.

Rosenberg, N.A. (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Schoville, S.D. & Roderick, G.K. (2010) Evolutionary diversification of cryophilic *Grylloblatta* species (Grylloblattodea: Grylloblattidae) in alpine habitats of California. *BMC Evolutionary Biology*, **10**, 163.

Schoville, S.D., Lam, A.W. & Roderick, G.K. (2012) A range-wide genetic bottleneck overwhelms contemporary landscape factors and local abundance in shaping genetic patterns of an alpine butterfly (Lepidoptera: Pieridae: *Colias behrii*). *Molecular Ecology*, **21**, 4242–4256.

Shafer, A.B.A., Cullingham, C.I., Côté, S.D. & Coltman, D.W. (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, **19**, 4589–4621.

Shapiro, A.M. (1975) Genetics, environment, and subspecies differences: the case of *Polites sabuleti* (Lepidoptera: Hesperiidae). *Great Basin Naturalist*, **35**, 33–38.

Shapiro, A.M. & Manolis, T.D. (2007) *Field guide to butterflies of the San Francisco Bay and Sacramento Valley regions*. University of California Press, Berkeley, CA.

U.S. Fish and Wildlife Service (2012) Endangered and threatened wildlife; 12-month finding on a petition to list four subspecies of Great Basin butterflies as endangered or threatened species. *Federal Register*, **77**, 54294–54329.

Vekemans, X. (2002) *AFLP-SURV version 1.0*. Distributed by the author. Laboratoire de Génétique et Écologie Végétale, Université Libre de Bruxelles, Belgium.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Paleman, J., Kuiper, M. & Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Warren, A.D., Davis, K.J., Stangeland, E.M., Pelham, J.P. & Grishin, N.V. (2012) *Illustrated lists of American butterflies*. Available at: http://www.butterfliesofamerica.com/.

Wilson, J.S. & Pitts, J.P. (2008) Revision of velvet ant genus *Dilophotopsis* Schuster (Hymenoptera: Mutillidae) by using molecular and morphological data, with implications for desert biogeography. *Annals of the Entomological Society of America*, **101**, 514–524.

Wilson, J.S. & Pitts, J.P. (2010a) Illuminating the lack of consensus among descriptions of earth history data in the North American deserts: a resource for biologists. *Progress in Physical Geography*, **34**, 419–441.

Wilson, J.S. & Pitts, J.P. (2010b) Phylogeographic analysis of the nocturnal velvet ant genus *Dilophotopsis* (Hymenoptera: Mutillidae) provides insights into diversification in the Nearctic deserts. *Biological Journal of the Linnean Society*, **101**, 360–375.

Wilson, J.S., Sneck, M., Murphy, D.D., Nice, C.C., Fordyce, J.A. & Forister, M.L. (2013a) Complex evolutionary history of the pallid dotted-blue butterfly (Lycaenidae: *Euphilotes pallescens*) in the Great Basin of western North America. *Journal of Biogeography*, **40**, 2059–2070.

Wilson, R.E., Peters, J.L. & McCracken, K.G. (2013b) Genetic and phenotypic divergence between low- and high-altitude populations of two recently diverged cinnamon teal subspecies. *Evolution*, **67**, 170–184.

Zink, R.M. (2004) The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 561–564.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Sampling data for individuals included in genetic analyses.

Appendix S2 Results from additional analyses (Figs S1–S6).

BIOSKETCH

Joshua P. Jahner is a PhD student at the University of Nevada, Reno, and his research interests focus on the influence of abiotic and biotic drivers of diversification for butterflies and moths found in western North America and the Neotropics.

Author contributions: J.P.J. led in writing the manuscript; J.P.J., M.L.F. and J.S.W. conducted the analyses; data were collected by M.L.F., C.C.N., J.A.F., J.S.W. and Z.H.M.; specimens were collected by M.L.F., C.C.N., J.A.F., D.D.M. and A.M.S.; the original idea for the phylogeography of this complicated, widespread, morphologically diverse skipper can be blamed on A.M.S.

Editor: Aristeidis Parmakelis