Environmental effects on gene flow in a species complex of vagile, hilltopping butterflies

JULIAN R. DUPUIS[1,2]*[,●], CATHERINE I. CULLINGHAM[2], SCOTT E. NIELSEN[3] and FELIX A. H. SPERLING[2]

[1]*Department of Plant and Environmental Protection Services, University of Hawai'i at Mānoa, Honolulu, HI 96822, USA*
[2]*Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada*
[3]*Department of Renewable Resources, University of Alberta, Edmonton, Alberta T6G 2H1, Canada*

Butterflies have played a pivotal role in our understanding of animal movement, but little is known about landscape-scale movement in highly vagile species with large ranges and open population structure. We investigate the effect of environment and landscape on both inter- and intraspecific genetic differentiation and population structure in the *Papilio machaon* group of swallowtail butterflies in Alberta, Canada. These butterflies exhibit strong hilltopping behaviour, where individuals congregate on topographical prominences to mate, and thus we expect topography to influence population connectivity. We use redundancy analysis to explore the relationship between two genetic datasets (mitochondrial DNA and microsatellites) and topographic, climatic and habitat-related environmental variables at multiple spatial scales. Shared variation across these environmental variables explained a large portion of both mitochondrial ($\leq 75\%$) and microsatellite genetic variation ($\leq 40\%$). Individual variables, such as elevation, terrain profile, solar radiation and precipitation, had moderate explanatory power; however, the water component of land cover had the greatest explanatory power of the individual variables and acted as a surrogate variable for the larval habitat of *P. machaon*. Despite high levels of retained ancestral polymorphism in the mitochondrial dataset, we found similar relationships between environmental characteristics and genetic variation. Finally, the spatial scale of our analyses greatly influenced our results, and without exploring the effect of scale we would have missed or discounted the effects of several environmental characteristics. This point emphasizes the importance of exploratory analysis in systems where there is little prior knowledge of movement dynamics and the influence of the environment.

ADDITIONAL KEYWORDS: landscape genetics – *Papilio machaon* – population genetics – spatial ecology.

INTRODUCTION

Butterflies have played a pivotal role in our understanding of metapopulation dynamics (Hanski & Thomas, 1994), the ecology of habitat fragmentation (Saccheri *et al.*, 1998; Hanski & Ovaskainen, 2000) and the influence of landscape features on individual dispersal and abundance (Roland *et al.*, 2000; Dover & Settele, 2009; Schultz *et al.*, 2019). However, much of this research has been conducted on species with a 'closed' population structure, i.e. those with low dispersal capability and small ranges (Dover & Settele, 2009). Many species of butterflies have high vagility and dispersal capability, and much less research has sought to characterize the influence of environmental and landscape features on highly vagile species. At the same time, many of these species exhibit behavioural characteristics that are clearly tied to landscape features and would be expected to have an effect on population structure and gene flow. 'Hilltopping' is a phenomenon in flying insects where adults congregate on topographical prominences (Shields, 1967 and references therein). Attraction to these geographical features is considered a strategy of mate-finding for species with low or patchy distributions (Shields, 1967; Alcock, 1987; Painter, 2014). Males populate hilltop locations awaiting females, and exhibit territorial behaviour in holding these locations (Scott, 1974;

*Corresponding author. E-mail: jrdupuis@hawaii.edu

Shields, 1967). Although 'topographical prominences' can vary in size, for butterflies they are most often the tops of hills, ridges, peaks and mountains (Shields, 1967; Alcock, 1987; Painter, 2014).

Much of the research on hilltopping dynamics has focused on measuring the movement of individuals with mark–recapture or similar approaches (Baughman & Murphy, 1988; Pe'er *et al.*, 2004; Grof-Tisza *et al.*, 2017), or using hilltopping as a model for simulating non-random dispersal (Pe'er *et al.*, 2005, 2006, 2013; Painter, 2014). Although these individual-focused approaches have clearly demonstrated the effect of elevation/topography as an orientation cue for hilltopping individuals (Pe'er *et al.*, 2004; Grof-Tisza *et al.*, 2017), less work has assessed the effect of these behaviours on population connectivity and gene flow. Additionally, although differential effects of landscape features at variable spatial scales are well known (e.g. Angelone *et al.*, 2011; Galpern *et al.*, 2012; Keller & Holderegger, 2013; Keller *et al.*, 2013), no work has been done to try to understand the effect of hilltopping on population dynamics at large geographical scales. Other environmental characteristics (temperature, precipitation, solar radiation, etc.) might be biologically relevant at larger landscape-level scales (e.g. Ramette & Tiedje, 2007; Hazard *et al.*, 2013) and might have an effect, or a combined effect with topography, on butterfly movement and population connectivity in these highly vagile species.

The Old World swallowtail butterfly, *Papilio machaon*, and anise swallowtail butterfly, *Papilio zelicaon*, occur in sympatry and parapatry across western North America, and morphologically intermediate hybrid individuals have been collected from central and southwest Alberta, Canada since the early 1900s (Scott, 1986; Sperling, 1987, 1990; Dupuis & Sperling, 2015). *Papilio machaon* is represented in Alberta by two regionally common subspecies: *Papilio machaon dodi* in the southern half of the province, and *Papilio machaon pikei* in the Peace River valley of northwestern Alberta and adjacent British Columbia (Sperling, 1987; Pelham, 2008). Both subspecies are restricted to arid river valley habitats, where their larvae feed on *Artemesia dracunculus* L. on eroding riverbanks, and adults hilltop on prominent edges of the river valleys (Scott, 1986; Sperling, 1987). The habitat and larval host plants of *P. zelicaon* are more diverse in western Canada, but populations are centred on isolated hilltops in boreal forest, prairies and foothills of the Rocky Mountains (Sperling, 1987; Bird *et al.*, 1995). Despite their ecological differences and broad sympatry, *P. machaon* and *P. zelicaon* appear to hybridize occasionally in southwest Alberta (Sperling, 1987, 1990; Sperling & Harrison, 1994; Dupuis & Sperling, 2015). Although their capacity for fast flight and potentially strong dispersal capability

is known (*P. zelicaon* in California has been marked and recaptured 4.3 km away in a single day; Shields, 1967), little is known of the dispersal dynamics of these species or how the surrounding environment affects their movement and population connectivity.

Here, we investigate the effect of landscape and environmental factors on gene flow and population structure between *P. machaon* and *P. zelicaon* in Alberta, Canada. We use mitochondrial DNA (mtDNA) and microsatellite datasets generated by Dupuis & Sperling (2016) and multiple forms of redundancy analysis (RDA) to assess the variation in environmental and landscape characteristics that explains the variation found in each genetic dataset. Using this information, we ask three questions. (1) Can landscape or environmental variables explain inter- and intraspecific genetic diversity and differentiation in this system? (2) Do landscape effects differ between nuclear and mitochondrial markers? Finally, (3) do landscape or environmental effects differ across spatial scales? To account for butterfly perception of land cover and topography, we use neighbourhood analysis on rasters of land cover, in addition to combined rasters of terrain and topography that model topographical prominence. Given that little is known of butterfly movement in this system, we conduct analyses at multiple geographical scales, through multiple neighbourhood analyses. This is the first use of such models of topographical prominence for modelling gene flow of a hilltopping insect and can guide future research in such systems.

GENETIC DATASETS

To provide context for the present analyses, we briefly summarize the population genetic datasets and results of Dupuis & Sperling (2016) as they apply to this study (see Fig. 1). A total of 822 specimens were collected from four regions in Alberta and adjacent northeastern British Columbia: Red Deer River valley, Peace River valley, Swan Hills, and Foothills of the Rocky Mountains. These collection localities were either hilltop locations, where we collected adults (generally a small area < 100 m in diameter on hills/mountains, or along a ridgeline/river valley bank for 200–300 m), or patches of host plant, where we collected larvae (generally small areas of 100–500 m², often along roadsides). Some collection localities consisted of both adult and larval collection where these areas were adjacent or within 1–2 km of each other.

Genotype data were collected for ten microsatellite loci (Zakharov & Hellmann, 2007) and an 831 bp region of mitochondrial cytochrome oxidase subunit I (*COI*). We conducted comprehensive phylogenetic and population genetic analyses to assess population structure, relatedness and putative hybridization
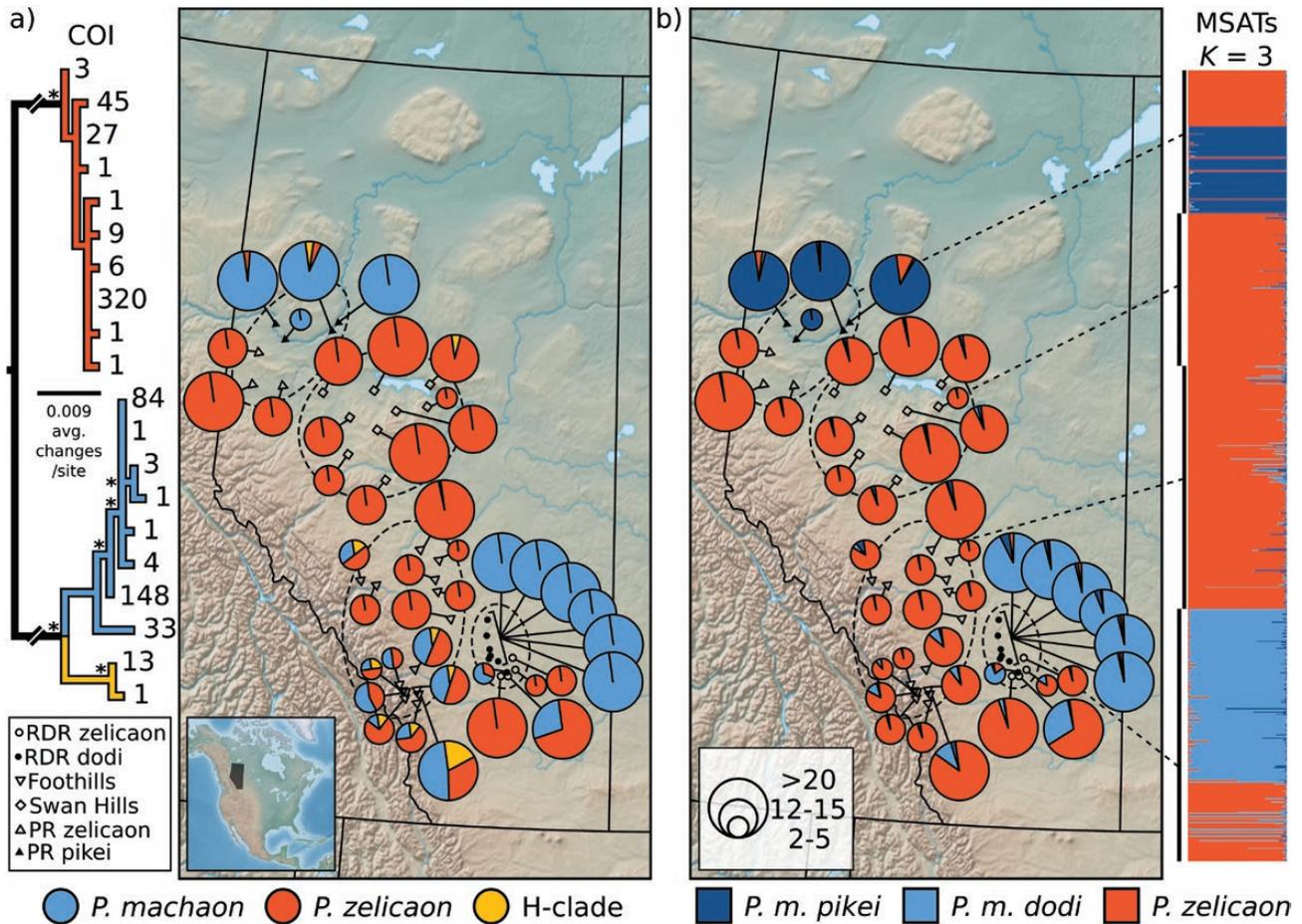
Figure 1. Summarized genetic results from Dupuis & Sperling (2016) for *COI* (A) and microsatellite (MSATs) datasets (B). In A, pie charts show the proportion of *Papilio machaon*, *Papilio zelicaon* and H-clade *COI* haplotypes per population, corresponding to the 50% majority rule consensus phylogeny of unique haplotypes constructed using Bayesian inference; * indicates high node support (> 0.80 posterior probability), and terminal branch labels represent sample sizes for each unique haplotype. In B, pie charts show average *Q*-values per population from *K* = 3 STRUCTURE results. In both A and B, the pie chart size indicates the population sample size, and population markers on the maps correspond to main geographical regions and species as in Figure 2. Abbreviations: PR, Peace River; RDR, Red Deer River. Map layers are from the public domain at www.naturalearthdata.com

in this system. With mtDNA, we found broad phylogenetic patterns matching that of the species complex as a whole (Dupuis & Sperling, 2015), with strong divergence between *P. machaon* and *P. zelicaon* and between *P. machaon* and the *P. machaon*-like H-clade (Fig. 1A). The latter is a group of haplotypes found in hybrid lineages of this species complex across North America and is thought to be a signature of hybridization with an ancestral lineage of *P. machaon* (Dupuis & Sperling, 2015). Microsatellite results were largely congruent with the broad results from *COI* (Fig. 1B). There was strong support for two and three genetic clusters, with the former separating *P. machaon* from *P. zelicaon* and the latter further separating *P. m. dodi* and *P. m. pikei*. Hybrids were apparent in some regions, particularly the Foothills

and the Red Deer River valley. Overall, individual nuclear ancestry and mtDNA clade of an individual were correlated, thus confirming the taxonomic identity of these main clusters. In general, across our sampling we found *P. machaon* predominantly at localities associated with the steep, arid river banks and *P. zelicaon* at the isolated prairie/forest hilltops, away from the river valleys.

MATERIAL AND METHODS

GENETIC DATA PREPARATION

Owing to limitations in contiguous environmental/ landscape data across province boundaries, we excluded the seven populations from British Columbia, leading

to a sample size of 703 specimens from 42 populations. For the microsatellite dataset, allele counts in each population were transformed to Hellinger distances (Legendre & Gallagher, 2001) using the decostand function in the vegan package v.2.2-0 (Oksanen *et al.*, 2014) in R v.3.1.1 (R Core Team, 2018). This transformation was developed for species abundance data and is appropriate for microsatellite data with high allelic diversity that may be missing many values (absent alleles) in individual populations (Cullingham *et al.*, 2014). For *COI*, we generated a matrix of pairwise Jost's *D* statistics (Jost, 2008) calculated from a FASTA formatted sequence alignment (594 bp in length, eliminating variable missing data from the 3′ and 5′ ends of the sequencing product) with the R packages APE v.5.1 (Paradis *et al.*, 2004) and MMOD v.1.3.3 (Winter, 2012).

ENVIRONMENTAL/LANDSCAPE CHARACTERISTICS

We used two types of environmental and landscape variables as predictors in the redundancy analyses: point estimates for collection localities, and local averages of raster-based data calculated with neighbourhood analysis. We generally chose variables that are either known to be important for insect survival/movement or that we hypothesized would have high explanatory power with regard to swallowtail butterfly movement. Data sources, with the rationale and general predictions for each variable type, are given in Table 1. For basic climate variables, we used point estimates for collection localities, generated with ClimateNA v.5.21 (Hamann *et al.*, 2013). For climate variables, we chose several measurements of a general variable type (e.g. for temperature-related variables, we included mean annual temperature, degree-days above 5 °C, extreme minimum temperature over 30 years, etc.) to provide optimal potential explanatory power (for details, see Table 1).

Given our knowledge of individual movement/ behaviour in this species group, we expect that characteristics of topography and land cover will be highly explanatory for measures of genetic diversity and relatedness. However, variables based on these characteristics are most biologically relevant when they are considered in the context of a local spatial environment; for instance, the topographic prominence of a hilltopping location is much better represented biologically when the surrounding landscape and relative elevations are taken into account, rather than when it is represented by a single measurement of elevation at the peak of the hilltop. Thus, for raster-based variables associated with topography and land cover (terrain position index, solar heatload index and land cover; Table 1), we used the Neighbourhood toolset in ArcGIS to generate average values representing the neighbourhood of raster cells around collection localities. To account for potential hierarchical spatial scales (e.g. Mullen *et al.*, 2010), we calculated variables using three different neighbourhood sizes based on the length from the centre raster cell (the cell containing the collection locality) to the edge of the neighbourhood in any direction: 1000, 250 and 100 m. Neighbourhoods were square in shape and centred on the centre raster cell. Given the variable resolution of these raster layers (land cover was a 100 m raster, whereas terrain position and solar heat load indices were 50 m rasters), the exact neighbourhood size differed between variables (see Supporting Information, Fig. S1), but for consistency we refer to these three neighbourhood sizes by the aforementioned centre-to-edge lengths. Although dispersal of up to 4.3 km is documented in *P. zelicaon* in California (Shields, 1967), we focused on smaller distances with the rationale that these would more effectively represent terrain and land-cover characteristics in Alberta.

To avoid problems associated with multicollinearity (Graham, 2003), we filtered the full set of environmental variables for each neighbourhood size by first removing variables correlated > 0.90. We then removed additional variables that met two criteria: (1) those that explained a relatively low amount of the variance in the genetic data (< 5%), assessed with variance partitioning (Borcard *et al.*, 1992; ter Braak & Verdonschot, 1995) conducted with the vegan package in R; and (2) those with a variance inflation factor > 10 (Zuur, Ieno & Elphick, 2010), calculated with the car library (Fox & Weisberg, 2011) in R. This approach resulted in sets of environmental variables with low multicollinearity (variance inflation factor < 10) and high explanatory power for the genetic data and retained more environmental variables than would have been retained with a lower initial correlation threshold. For the final set of variables, we also used the *stepAIC* function in the MASS library (Venables & Ripley, 2002) in R to perform additional model selection and ensure that the change in the Akaike information criterion was less than two for the addition of any variable in the final model (Burnham & Anderson, 2002; Mazerolle, 2006).

REDUNDANCY ANALYSIS

To test what geographical and environmental characteristics might drive gene flow between populations, we conducted RDA (Rao, 1964). This method is a multivariate extension of multiple linear regression (Legendre & Legendre, 2012) and explores variation in a set of predictor variables that explains the variation present in a set of response variables (Florentino *et al.*, 2008; Vernesi *et al.*, 2012; Stevens *et al.*, 2016; Strobel *et al.*, 2016). In our case, the

Table 1. Environmental and landscape characteristics used in this study, their rationale, and broad predictions for major variable types

Variable	Description	Variable category	Source	Rationale and predictions
ELEV	Elevation (in metres)	Topography	NOAA NCEI GLOBE DEM: http://www.ngdc.noaa.gov/mgg/topo/globeget.html	Topological prominences are used as congregation and mating locations in hilltopping butterflies (Shields, 1967; Painter, 2014), so we expect these butterflies to be drawn to higher elevation and positive TPI values. (Positive TPI values indicate ridgetops and hilltops, whereas negative TPI indicates valleys and canyon bottoms, and flat areas have a TPI of zero.)
TPI	Terrain position index	Topography	Code from Jenness (2006)	
SOLAR	Solar heatload index	Topography	Equations from McCune (2007)	
MAT	Mean annual temperature (in degrees Celsius)	Climate	ClimateNA v.5.21 (Hamann et al., 2013)	As ectotherms, solar input and temperature play a major role in insect development, distribution and behaviour (Bale, 1991). Extreme cold temperatures affect overwintering survival (e.g. Marshall et al., 2015), and warm temperatures and degree-day accumulation have various
MWMT	Mean warmest month temperature (in degrees Celsius)	Climate	ClimateNA v5..21	
DD5	Degree-days above 5°C	Climate	ClimateNA v.5.21	effects on growth and development (e.g. Wilson & Barnett, 1983; Wallner, 1987; Gilbert & Rawthorn, 1996). We expect warm temperatures to facilitate population growth and butterfly
EMT	Extreme minimum temperature (in degrees Celsius) over 30 years	Climate	ClimateNA v.5.21	
SHM	Summer heat:moisture index	Climate	ClimateNA v.5.21	dispersal and extreme cold temperatures to do the opposite.
MAP	Mean annual precipitation (in millimetres)	Climate	ClimateNA v.5.21	Precipitation has wide-ranging effects on herbivorous insect population dynamics, largely
MSP	Mean summer (May to September) precipitation (in millimetres)	Climate	ClimateNA v.5.21	through effects of drought on host plants (Ehrlich et al., 1972; Morecroft et al., 2002; Raimondo et al., 2004; Robinson et al., 2012). We expect precipitation to facilitate survival and population growth, and thus lead to increased emigration and dispersal.
LC	Land cover	Habitat	National Resource Canada, Earth Observation for Sustainable Development of Forests, Landcover Classification Scheme 2003: http://www.nrcan.gc.ca/forests/measuring-reporting/remote-sensing/13433	These species of swallowtail butterflies are generally open-habitat species, avoiding closed and heavily forested habitats (Sperling, 1987); therefore, we expect open habitat types to facilitate butterfly dispersal. Classifications: broad, broadleaf forest; conif, coniferous forest; herb, herbaceous/non-woody vegetation; mixed, mixed wood forest; nonveg, non-vegetated; shrub, shrubland; water, lake/pond/river; wetland, wetland/marsh.

Acronyms are defined in the Description column.

population-based genetic data were the response variable, and the predictors were environmental characteristics at and around each locality. For the microsatellite dataset, Hellinger distances (based on allele counts) were used as the response variable, with the standard *rda* function in the vegan package v.2.2-0 (Oksanen *et al.*, 2014) in R v.3.1.1 (R Core Team, 2018). For *COI*, a matrix of pairwise Jost's *D* statistics were used as the response variable for distance-based redundancy analysis (dbRDA; Legendre & Anderson, 1999) using the *capscale* function in vegan v.2.2-0.

We also wanted to distinguish between the individual variability explained by each environmental variable and any shared contribution of multiple variables explaining genetic variation. To do so, we conducted variance partitioning on three main categories of environmental/landscape variables: topography, climate and habitat (Table 1). This method separates the total variation explained in the genetic data into variation explained by individual variable categories and combinations of variable categories. To assess the significance of the variable categories, we conducted additional RDA/dbRDA on categories individually and used ANOVA in R with 999 permutations. We conducted separate RDA/dbRDA (in addition to filtering for multicollinearity and variance partitioning) for each neighbourhood size of environmental/landscape data, leading to six sets of analyses (microsatellite and *COI* data for each of the 1000, 250 and 100 m neighbourhoods).
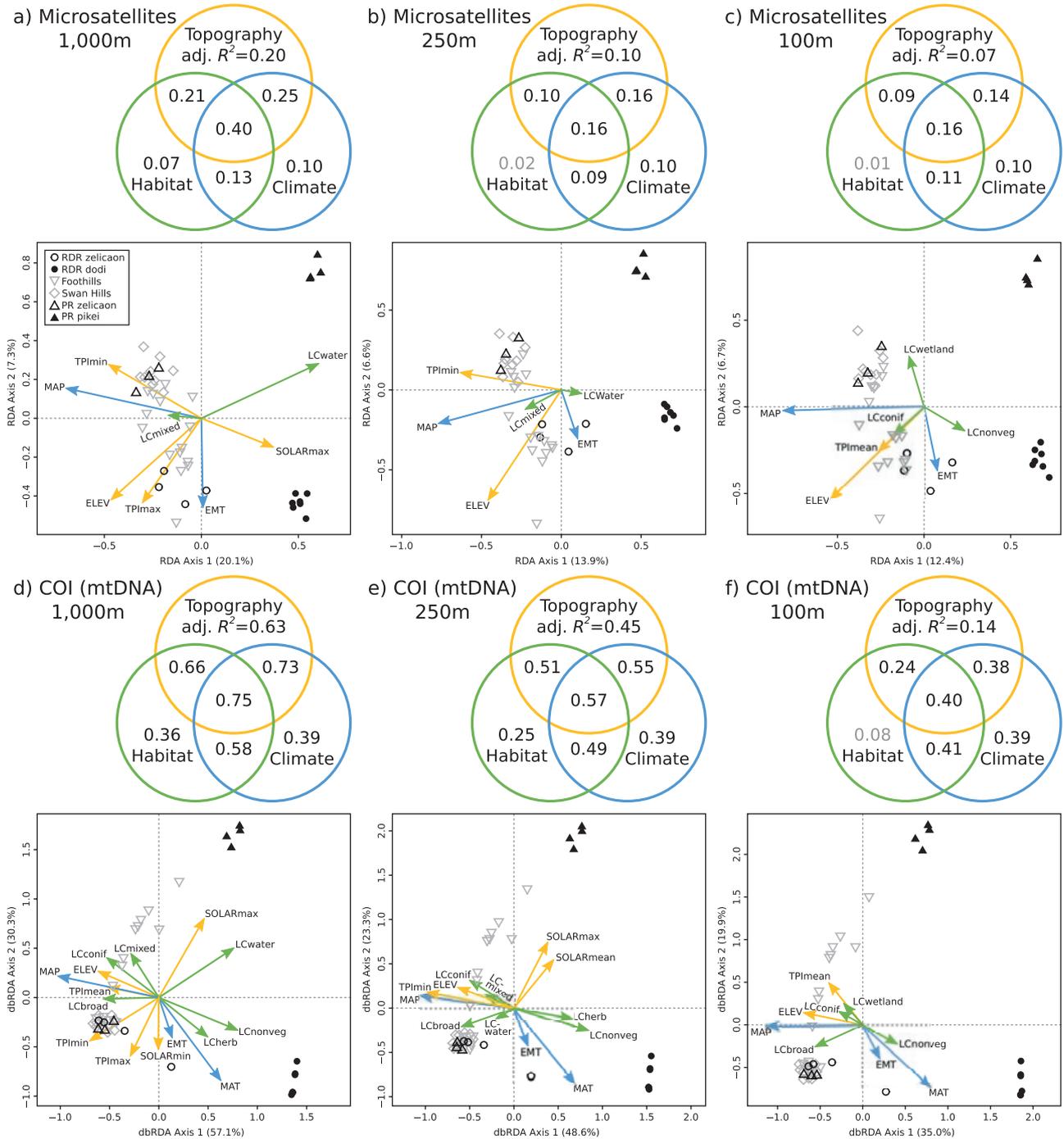
RESULTS

REDUNDANCY ANALYSIS

In all analyses, the first RDA axis separated *P. machaon* from *P. zelicaon*, and the second axis separated the subspecies of *P. machaon* and population divisions in *P. zelicaon*. The final set of variables included as predictors in the RDAs (microsatellite dataset) differed from those in the dbRDAs (*COI* dataset); however, the predictor variables were relatively consistent between different neighbourhood sizes for both the RDA and the dbRDA (Fig. 2). Overall, more variables were included in the final variable sets for *COI*, compared with the microsatellites. In all of the analyses, elevation (ELEV) was correlated with mean warmest month temperature (MWMT) and degree-days above 5°C (DD5), and mean annual precipitation (MAP) was correlated with mean summer precipitation (MSP) and summer heat:moisture index (SHM).

Generally, all three categories of variables (topographic, climatic and habitat based) independently explained significant proportions of both microsatellite allelic variation and *COI* distance variation (Fig. 2; non-shared parts of Venn diagrams); the only exception to this trend was that habitat-based variables did not explain a significant proportion of genetic variation at the smallest geographical scale (100 m neighbourhoods) for both microsatellites and *COI*, and the mid-range scale for microsatellites. Despite the independent explanatory power of each variable category, variance partitioning identified a large proportion of genetic variation that was explained by shared variation between variable categories in all analyses (Fig. 2; shared parts of Venn diagrams). At larger geographical scales (1000 and 250 m neighbourhoods), the topographic variable category (elevation and solar radiation and terrain position indices) had higher explanatory power (when considering both independent and shared measurements) than climatic and habitat-based variable categories. However, at smaller scales (100 m neighbourhoods), climatic variables had higher explanatory power.

In ordination space, microsatellite allelic variation between *P. machaon* and *P. zelicaon* was best explained by a combination of maximum solar radiation (SOLARmax) and land cover consisting of water (LCwater) (Fig. 2A). At smaller scales, the explanatory power for the large genetic differentiation between species was weaker, but still explained by land-cover variables (Fig. 2B, C). Among *P. zelicaon* populations, MAP and ELEV explained genetic differentiation consistently across geographical scales (Fig. 2A–C). Microsatellite allelic variation between predominantly *P. zelicaon* populations ('RDR zelicaon', 'PR zelicaon' and the Swan Hills and Foothills) was less well defined, although some general patterns were apparent. For example, general north-to-south allelic variation was apparent and was relatively consistently explained by a combination of ELEV, terrain profile index (TPImin/mean/max) and extreme minimum temperature (EMT) across all geographical scales (Fig. 2A–C).

For mtDNA, the genetic divergence between *P. machaon* and *P. zelicaon* was visually explained by a large combination of variables, including solar radiation (SOLARmin/mean/max), mean annual temperature (MAT) and various land-cover characteristics (Fig. 2D–F). As with the microsatellite dataset, MAP and ELEV consistently explained *P. zelicaon* variation across all spatial scales (Fig. 2D–F), and LCwater was explanatory for *P. machaon* only at the largest geographical scale (Fig. 2D). There was little geographical structure between regions among the predominantly *P. zelicaon* populations with mtDNA. The only discernable structure in *COI* genetic distance in these predominantly *P. zelicaon* populations was the separation of some of the Foothills region populations, owing to the presence *P. machaon* and H-clade haplotypes in putative hybrid individuals (Fig. 1). Maximum solar radiation (SOLARmax)

Figure 2. Results of variance partitioning and redundancy analyses (RDA) of Hellinger-transformed allele counts from microsatellite data (A–C), and variance partitioning and distance-based redundancy analysis (dbRDA) of Jost's *D* genetic distance matrix from *COI* data (D–F). For each panel, the Venn diagram contains adjusted R^2 values for the separate and combined effects of colour-coded variable categories (topography, climate and habitat) calculated with variance partitioning; grey text indicates non-significant relationships ($P > 0.05$). Each panel of A–C and D–F correspond to different neighbourhood sizes in the calculation of environmental and landscape variables. Some clusters of populations in D–F have been separated manually to clarify their placement. The prefix LC indicates land-cover variables. Topography variables: ELEV, elevation; SOLARmin/mean/max, solar radiation index; TPImin/mean/max, terrain position index. Climate variables: EMT, extreme minimum temperature; MAP, mean annual precipitation; MAT, mean annual temperature. Habitat variables: LCbroad, broadleaf forest; LCconif, coniferous forest; LCherb, herbaceous vegetation; LCmixed, mixed forest; LCnonveg, non-vegetated/developed; LCwetland, wetland.

is likely to hold some explanatory power for this differentiation, at least at larger geographical scales (1000 and 250 m; Fig. 2D, E).

DISCUSSION

Do environmental variables explain genetic differentiation?

We identified a combination of environmental and landscape variables that explained genetic differentiation between *P. machaon* and *P. zelicaon*. The presence of water as land cover (LCwater) had high explanatory power with both genetic datasets, which might seem counterintuitive at first, because these species show no particular affinity for open water. However, rivers and their river valleys represent substantial features in these regions. Around the collection localities, the Red Deer River and Peace River are ~100–200 and ~200–1500 m wide, respectively, and the river valleys are ~1.5–3.5 and ~2.5–5.5 km from edge to edge, respectively. Collection localities for *P. machaon* are either within the river valley itself (larval collections on host plant) or at the top edge of the river valley (adult collections at hilltopping locations). Thus, in these analyses, proximity to water acts as a surrogate variable for the arid river valley habitat used by *P. machaon* and its host *A. dracunculus* (Sperling, 1987). Host occurrence has been shown to have primary roles in butterfly movement and population dynamics (Hanski & Heino, 2003; Fred *et al.*, 2006). We expect that a variable indicating host plant occurrence would also be highly explanatory for species distributions and diversity (e.g. Wiklund *et al.*, 2018). Unfortunately, accurate, high-resolution host plant data on a provincial scale were unavailable for this study.

In general, we expected topographic prominences and open land-cover classes to facilitate gene flow given the general behaviour of these species (Dupuis & Sperling, 2015) and previous findings in related butterfly species (Roland *et al.*, 2000; Keyghobadi *et al.*, 2005). We found that ELEV and the terrain profile and solar radiation indices (TPI and SOLAR) were often highly explanatory, and probably contributed to the overall high explanatory power of the topographic variable category. Additionally, open land-cover classes (e.g. LCnonveg, LCherb) were also explanatory, particularly at smaller geographical scales, despite the overall lower, and often non-significant, explanatory power of the habitat-related variable category. The larger relative effect of these open land-cover classes at smaller geographical scales makes sense ecologically, because the hilltop locations in more forested regions (the Swan Hills and Foothills) are characterized by large open areas at the peaks of hilltops and mountains. Thus, overall our expectations for these landscape features were confirmed, although more even sampling would strengthen this conclusion (as discussed below).

Generally, shared variation between topographic, climatic and habitat-related variables explained the variation in genetic datasets better than independent variation in each variable category. This shared explanatory power seems to be important when considering intraspecific differentiation in the predominantly *P. zelicaon* populations, because no single variable clearly explained that axis of variation. Although explaining the intraspecific differentiation in *P. zelicaon* populations is inherently limited in this dataset, given the lower differentiation between these populations (i.e. there is simply less genetic differentiation to explain), many combinations of variables seem reasonable given the characteristics of these geographical regions.

Although we focused here on exogenous factors affecting genetic differentiation, there are clearly also endogenous factors at play. Some of these factors are relatively simple, such as isolation-by-distance explaining genetic differentiation between *P. m. dodi* and *P. m. pikei*. Others are more complex, such as potential genetic compatibility. Although any of these populations can hybridize (e.g. Clarke & Sheppard, 1955; Ae, 1979), and occasionally do in nature (Dupuis & Sperling, 2015, 2016), it is still unclear how contemporary hybridization and ancestral introgression combine to create the biological reality observed today.

Interestingly, Sperling (1987) concluded that mean annual temperature (based on records from 1951 to 1980) better explained the distributions of many taxonomic entities in this species group (particularly subspecies of *P. machaon* in Alberta and British Columbia), and that mean annual precipitation was more homogeneous for the species and subspecies in this region. We found that extreme minimum temperature and mean annual temperature were relatively explanatory in our analysis, but that climatic variables were overall less broadly explanatory than other variables. This apparent contradiction between studies could indicate temporal changes in these variables (the temperature-based data here are primarily derived from the period from 1961 to 1990) or differences attributable to methodology. However, the more restricted scope of analysis in the present study is likely to play a larger role, because we were concerned with a geographically much smaller range of the systematic diversity than that considered by Sperling (1987).

Do landscape effects differ between genetic markers/spatial scales?

Overall, some similar trends were apparent from both genetic datasets; for example, LCwater/SOLARmax

and ELEV/MAP explained differentiation between species for both *COI* and microsatellites. This was despite the inherent differences between the marker types, notably the different evolutionary timescales of mtDNA and microsatellites, in addition to the large amount of retained ancestral polymorphism present in the *COI* dataset for this group (Dupuis & Sperling, 2015, 2016). In contrast, the microsatellite dataset clearly differentiates *P. m. dodi* and *P. m. pikei* and provides more geographical separation of the *P. zelicaon* populations (Fig. 2). Owing to this pattern across the marker sets, and the more strictly neutral nature of microsatellites (compared with mtDNA; Ballard & Kreitman, 1995; Ballard & Whitlock, 2004), we have more confidence in the results from the microsatellite dataset. However, the concordance of results between genetic datasets increases our confidence in the effects of these landscape characteristics on genetic differentiation in this system.

We found that the spatial scale in the neighbourhood analysis greatly influenced the explanatory power of environmental variables. This is perhaps most clearly exemplified by the water land-cover variable, which acted as a surrogate variable for the habitat of *P. machaon*. Although this surrogate variable had high explanatory power for both datasets, it had this effect in the analyses only at the largest geographical scale. If only smaller geographical scales had been explored here, we would have missed these patterns. Another example of the importance of spatial scale is the general shift from topographic variables having higher relative explanatory power at larger geographical scales to climatic variables having higher relative power at the smallest geographical scale. Many studies have identified landscape features having different effects at varied spatial scales or variable landscape effects for short- vs. long-distance dispersal (e.g. Angelone *et al.*, 2011; Galpern *et al.*, 2012; Keller & Holderegger, 2013; Keller *et al.*, 2013). That might be the case here, where topographic features (topography, elevation) are affecting more long-range movements, whereas shorter-range movements (in areas with higher density populations) are affected more by climatic variables; specifically, differences in precipitation (and probably more importantly, the lack of precipitation, i.e. drought) and temperature can affect insects both directly (particularly larval stages with lower vagility) and indirectly through their effects on host plant condition and survival (Ehrlich *et al.*, 1972; Wallner, 1987; Morecroft *et al.*, 2002; Hanski & Heino, 2003; Raimondo *et al.*, 2004; Fred *et al.*, 2006; Robinson *et al.*, 2012; Oliver *et al.*, 2015). The main limitation in extending these findings to the *P. machaon* system is our lack of evenly distributed sampling effort (a common limitation in many landscape genetic studies; Schwartz & McKelvey, 2009 and references therein); therefore, our conclusions regarding short- vs. long-distance dispersal remain preliminary. Regardless, these results highlight the importance of testing environmental effects at multiple geographical scales, which is amplified in systems such as this, where little is known of general movement dynamics.

CONCLUSIONS

Population genetic analyses of these data have suggested that the distributions of these highly vagile, hilltopping swallowtail butterflies are limited by both environment and landscape factors (Dupuis & Sperling, 2016), implying that gene flow between these species is limited by these factors. To test this, we asked what components of habitat explained gene flow both within and between the species. We identified environmental variables that explained significant genetic variation, and these variables differed across scales, suggesting that different processes (large-scale dispersal vs. local movements) are influenced by different factors. The detection of influences of landscape features depended on uniformity of habitat connectivity and contrast between highly resistant and optimal landscapes (Short Bull *et al.*, 2011; Cushman *et al.*, 2013), in addition to the distribution of sampled populations (Storfer *et al.*, 2007; Schwartz & McKelvey, 2009). Although our results might have been affected by the low power of analyses to detect landscape genetic patterns (Cushman & Landguth, 2010) or bias owing to incomplete sampling (Schwartz & McKelvey, 2009), these results give a strong starting point for considering the effect of landscape and environmental features on movement and gene flow in this system. Future research with more even sampling of populations at various spatial scales should increase insight into the effects of environmental and landscape characteristics in this and other highly vagile insect species.

REFERENCES

Ae SA. 1979. The phylogeny of some *Papilio* species based on interspecific hybridization data. *Systematic Entomology* **4:** 1–16.

Alcock J. 1987. Leks and hilltopping in insects. *Journal of Natural History* **21:** 319–328.

Angelone S, **Kienast F**, **Holderegger R. 2011.** Where movement happens: scale-dependent landscape effects on genetic differentiation in the European tree frog. *Ecography* **34:** 714–722.

Bale JS. 1991. Insects at low temperature: a predictable relationship? *Functional Ecology* **5:** 291–298.

Ballard JW, **Kreitman M. 1995.** Is mitochondrial DNA a strictly neutral marker? *Trends in Ecology & Evolution* **10:** 485–488.

Ballard JW, **Whitlock MC. 2004.** The incomplete natural history of mitochondria. *Molecular Ecology* **13:** 729–744.

Baughman JF, **Murphy DD. 1988.** What constitutes a hill to a hilltopping butterfly? *American Midland Naturalist* **120:** 441–443.

Bird CD, **Hilchie GJ**, **Kondla NG**, **Pike EM**, **Sperling FAH. 1995.** *Alberta butterflies*. Edmonton: Provincial Museum of Alberta.

Borcard D, **Legendre P**, **Drapeau P. 1992.** Partialling out the spatial component of ecological variation. *Ecology* **73:** 1045–1055.

ter Braak CJF, **Verdonschot PFM. 1995.** Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic Sciences* **57:** 255–289.

Burnham KP, **Anderson DR. 2002.** *Model selection and multimodel inference: a practical information-theoretic approach*, *2nd edn*. New York: Springer.

Clarke CA, **Sheppard PM. 1955.** A preliminary report on the genetics of the machaon group of swallowtail butterflies. *Evolution* **9:** 182–201.

Cullingham CI, **Cooke JE**, **Coltman DW. 2014.** Cross-species outlier detection reveals different evolutionary pressures between sister species. *The New Phytologist* **204:** 215–229.

Cushman SA, **Landguth EL. 2010.** Scale dependent inference in landscape genetics. *Landscape Ecology* **25:** 967–979.

Cushman S, **Shirk AJ**, **Landguth EL. 2013.** Landscape genetics and limiting factors. *Conservation Genetics* **14:** 263–274.

Dover J, **Settele J. 2009.** The influences of landscape structure on butterfly distribution and movement: a review. *Journal of Insect Conservation* **13:** 3–27.

Dupuis JR, **Sperling FA. 2015.** Repeated reticulate evolution in North American *Papilio machaon* group swallowtail butterflies. *PLoS One* **10:** e0141882.

Dupuis JR, **Sperling FA. 2016.** Hybrid dynamics in a species group of swallowtail butterflies. *Journal of Evolutionary Biology* **29:** 1932–1951.

Ehrlich PR, **Breedlove DE**, **Brussard PF**, **Sharp MA. 1972.** Weather and the "regulation" of subalpine populations. *Ecology* **53:** 243–247.

Florentino V, **Manganelli G**, **Giusti F. 2008.** Multiple scale patterns of shell and anatomy variability in land snails: the case of the Sicilian *Marmorana* (Gastropoda: Pulmonata, Helicidae). *Biological Journal of the Linnean Society* **93:** 359–370.

Fox J, **Weisberg S. 2011.** *An {R} companion to applied regression*, *2nd edn*. Thousand Oaks: Sage Publishing.

Fred MS, **O'Hara RB**, **Brommer JE. 2006.** Consequences of the spatial configuration of resources for the distribution and dynamics of the endangered *Parnassius apollo* butterfly. *Biological Conservation* **130:** 183–192.

Galpern P, **Manseau M**, **Wilson P. 2012.** Grains of connectivity: analysis at multiple spatial scales in landscape genetics. *Molecular Ecology* **21:** 3996–4009.

Gilbert N, **Rawthorn DA. 1996.** Forum: insects and temperature—a general theory. *The Canadian Entomologist* **128:** 1–13.

Graham MH. 2003. Confronting multicollinearity in ecological multiple regression. *Ecology* **84:** 2809–2815.

Grof-Tisza P, **Steel Z**, **Cole EM**, **Holyoak M**, **Karban R. 2017.** Testing predictions of movement behaviour in a hilltopping moth. *Animal Behaviour* **133:** 161–168.

Hamann A, **Wang T**, **Spittlehouse DL**, **Murdock TQ. 2013.** A comprehensive, high-resolution database of historical and projected climate surfaces for Western North America. *Bulletin of the American Meteorological Society* **94:** 1307–1309.

Hanski I, **Heino M. 2003.** Metapopulation-level adaptation of insect host plant preference and extinction–colonization dynamics in heterogeneous landscapes. *Theoretical Population Biology* **64:** 281–290.

Hanski I, **Ovaskainen O. 2000.** The metapopulation capacity of a fragmented landscape. *Nature* **404:** 755–758.

Hanski I, **Thomas CD. 1994.** Metapopulation dynamics and conservation: a spatially explicit model applied to butterflies. *Biological Conservation* **68:** 167–180.

Hazard C, **Gosling P**, **van der Gast CJ**, **Mitchell DT**, **Doohan FM**, **Bending GD. 2013.** The role of local environment and geographical distance in determining community composition of arbuscular mycorrhizal fungi at the landscape scale. *The ISME Journal* **7:** 498–508.

Jenness J. 2006. *Topographic position index (tip_jen.avx) extension for ArcView 3.x, v1.2.* Available at http://www.jennessent.com/arcview/tpi.htm

Jost L. 2008. G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* **17:** 4015–4026.

Keller D, **Holderegger R. 2013.** Damselflies use different movement strategies for short- and long-distance dispersal. *Insect Conservation and Diversity* **6:** 590–597.

Keller D, **Holderegger R**, **van Strien MJ. 2013.** Spatial scale affects landscape genetic analysis of a wetland grasshopper. *Molecular Ecology* **22:** 2467–2482.

Keyghobadi N, **Roland J**, **Strobeck C. 2005.** Genetic differentiation and gene flow among populations of the alpine butterfly, *Parnassius smintheus*, vary with landscape connectivity. *Molecular Ecology* **14:** 1897–1909.

Legendre P, **Anderson MJ. 1999.** Distance-based redundancy analysis: testing multspecies responses in multifactorial ecological experiments. *Ecological Monographs* **69:** 1–24.

Legendre P, **Gallagher ED. 2001.** Ecologically meaningful transformations for ordination of species data. *Oecologia* **129:** 271–280.

Legendre P, **Legendre L. 2012.** *Numerical ecology*, *3rd edn*. Amsterdam: Elsevier.

Marshall KE, **Sinclair BJ**, **Lee KP. 2015.** The relative importance of number, duration and intensity of cold stress events in determining survival and energetics of an overwintering insect. *Functional Ecology* **29:** 357–366.

Mazerolle MJ. 2006. Improving data analysis in herpetology: using Akaike's Information Criterion (AIC) to assess the strength of biological hypotheses. *Amphibia-Reptilia* **27:** 169–180.

McCune B. 2007. Improved estimates of incident radiation and heat load using non-parametric regression against topographic variables. *Journal of Vegetation Science* **18:** 751–754.

Morecroft MD, **Bealey CE**, **Howells O**, **Rennie S**, **Woiwod IP. 2002.** Effects of drought on contrasting insect and plant species in the UK in the mid-1990s. *Global Ecology and Biogeography* **11:** 7–22.

Mullen LB, **Woods HA**, **Schwartz MK**, **Sepulveda AJ**, **Lowe WH. 2010.** Scale-dependent genetic structure of the Idaho giant salamander (*Dicamptodon aterrimus*) in stream networks. *Molecular Ecology* **19:** 898–909.

Oksanen J, **Blanchet FG**, **Kindt R**, **Legendre P**, **Minchin PR**, **O'Hara RB**, **Simpson GL**, **Solymos P**, **Stevens HH**, **Wagner H. 2014.** *Vegan: community ecology package. R package v.2.2-0*. Available at http://CRAN.R-project.org/package=vegan

Oliver TH, **Marshall HH**, **Morecroft MD**, **Brereton T**, **Prudhomme C**, **Huntingford C. 2015.** Interacting effects of climate change and habitat fragmentation on drought-sensitive butterflies. *Nature Climate Change* **5:** 941–945.

Painter KJ. 2014. Multiscale models for movement in oriented environments and their application to hilltopping in butterflies. *Theoretical Ecology* **7:** 53–75.

Paradis E, **Claude J**, **Strimmer K. 2004.** APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20:** 289–290.

Pe'er G, **Heinz SK**, **Frank K. 2006.** Connectivity in heterogeneous landscapes: analyzing the effect of topography. *Landscape Ecology* **21:** 47–61.

Pe'er G, **Saltz D**, **Frank K. 2005.** Virtual corridors for conservation management. *Conservation Biology* **19:** 1997–2003.

Pe'er G, **Saltz D**, **Münkemüller T**, **Matsinos YG**, **Thulke H-H. 2013.** Simple rules for complex landscapes: the case of hilltopping movements and topography. *Oikos* **122:** 1483–1495.

Pe'er G, **Saltz D**, **Thulke H-H**, **Motro U. 2004.** Response to topography in a hilltopping butterfly and implications for modelling nonrandom dispersal. *Animal Behaviour* **68:** 825–839.

Pelham JP. 2008. A catalogue of the butterflies of the United States and Canada, with a complete bibliography of the descriptive and systematic literature. *The Journal of Research on the Lepidoptera* **40:** 1–672.

Raimondo S, **Liebhold AM**, **Strazanac JS**, **Butler L. 2004.** Population synchrony within and among Lepidoptera species in relation to weather, phylogeny, and larval phenology. *Ecological Entomology* **29:** 96–105.

Ramette A, **Tiedje JM. 2007.** Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 2761–2766.

Rao CR. 1964. The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A (1961–2002)* **26:** 329–358.

R Core Team. **2018.** *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/

Robinson N, **Armstead S**, **Deane Bowers M. 2012.** Butterfly community ecology: the influences of habitat type, weather patterns, and dominant species in a temperate ecosystem. *Entomologia Experimentalis et Applicata* **145:** 50–61.

Roland J, **Keyghobadi N**, **Fownes S. 2000.** Alpine *Parnassius* butterfly dispersal: effects of landscape and population size. *Ecology* **81:** 1642–1653.

Saccheri I, **Kuussaari M**, **Kankare M**, **Vikman P**, **Fortelius W**, **Hanski I. 1998.** Inbreeding and extinction in a butterfly metapopulation. *Nature* **392:** 491–494.

Schultz CB, **Haddad NM**, **Henry EH**, **Crone EE. 2019.** Movement and demography of at-risk butterflies: building blocks for conservation. *Annual Review of Entomology* **64:** 9.1–9.18.

Schwartz MK, **McKelvey KS. 2009.** Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics* **10:** 441–452.

Scott JA. 1974. Mate-locating behavior of butterflies. *American Midland Naturalist* **91:** 103–117.

Scott JA. 1986. *The Butterflies of North America: a natural history and field guide*. Stanford: Stanford University Press.

Shields O. 1967. Hilltopping: an ecological study of summit congregation behavior of butterflies on a southern California hill. *Journal of Research on the Lepidoptera* **6:** 69–178.

Short Bull RA, **Cushman SA**, **Mace R**, **Chilton T**, **Kendall KC**, **Landguth EL**, **Schwartz MK**, **McKelvey K**, **Allendorf FW**, **Luikart G. 2011.** Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Molecular Ecology* **20:** 1092–1107.

Sperling FAH. 1987. Evolution of the *Papilio machaon* species group in Western Canada (Lepidoptera: Papilionidae). *Quaestiones Entomologicae* **23:** 198–315.

Sperling FAH. 1990. Natural hybrids of *Papilio* (Insecta: Lepidoptera): poor taxonomy or interesting evolutionary problem? *Canadian Journal of Zoology* **68:** 1790–1799.

Sperling FAH, **Harrison RG. 1994.** Mitochondrial DNA variation within and between species of the *Papilio machaon* group of swallowtail butterflies. *Evolution; International Journal of Organic Evolution* **48:** 408–422.

Stevens RD, **Johnson ME**, **McCulloch ES. 2016.** Geographic variation of wing morphology of great fruit-eating bats (*Artibeus lituratus*): environmental, genetic and spatial correlates of phenotypic differences. *Biological Journal of the Linnean Society* **118:** 734–744.

Storfer A, **Murphy MA**, **Evans JS**, **Goldberg CS**, **Robinson S**, **Spear SF**, **Dezzani R**, **Delmelle E**, **Vierling L**, **Waits LP. 2007.** Putting the "landscape" in landscape genetics. *Heredity* **98:** 128–142.

Strobel HM, **Alda F**, **Sprehn CG**, **Blum MJ**, **Heins DC. 2016.** Geographic and host-mediated population genetic structure in a cestode parasite of the three-spined stickleback. *Biological Journal of the Linnean Society* **119:** 381–396.

The Inkscape Team. **2017.** *Inkscape v0.91.* Available at: https://inkscape.org

Venables WN, **Ripley BD. 2002.** *Modern applied statistics with S, 4th edn.* New York: Springer.

Vernesi C, **Rocchini D**, **Pecchioli E**, **Neteler M**, **Verndramin GG**, **Paffetti D. 2012.** A landscape genetic approach reveals ecological-based differentiation in populations of holm oak (*Quercus ilex* L.) at the northern limit of its range. *Biological Journal of the Linnean Society* **107:** 458–467.

Wallner WE. 1987. Factors affecting insect population dynamics: differences between outbreak and non-outbreak species. *Annual Review of Entomology* **32:** 317–340.

Wiklund C, **Norén K**, **Ryman N**, **Friberg M. 2018.** Local monophagy and between-site diversity in host use in the European swallowtail butterfly, *Papilio machaon*. *Biological Journal of the Linnean Society* **123:** 179–190.

Wilson LT, **Barnett WW. 1983.** Degree-days: an aid in crop and pest management. *California Agriculture* **37:** 4–7.

Winter DJ. 2012. MMOD: an R library for the calculation of population differentiation statistics. *Molecular Ecology Resources* **12:** 1158–1160.

Zakharov EV, **Hellmann JJ. 2007.** Characterization of 17 polymorphic microsatellite loci in the Anise swallowtail, *Papilio zelicaon* (Lepidoptera: Papilionidae), and their amplification in related species. *Molecular Ecology Notes* **7:** 144–146.

Zuur AF, **Ieno EN**, **Elphick CS. 2010.** A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* **1:** 3–14.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Graphical depiction of the neighbourhood sizes for 50 and 100 m rasters.