Considerable genetic diversity and structure despite narrow endemism and limited ecological specialization in the Hayden's ringlet, *Coenonympha haydenii*

Amy Springer[1], **Zachariah Gompert**[1,2]

[1] Department of Biology, Utah State University, Logan, UT 84322, USA

[2] Ecology Center, Utah State University, Logan, UT 84322, USA

Send correspondence to:
amy.mo.springer@gmail.com and zach.gompert@usu.edu

Abstract

Understanding he processes that underlie the development of population genetic structure is central to the study of evolution. Patterns of genetic structure, in turn, can reveal signatures of isolation by distance, barriers to gene flow, or even the genesis of speciation. However, it is unclear how severe range restriction might impact the processes that dominate the development of genetic structure. In narrow endemic species, is population structure likely to be adaptive in nature, or rather the result of genetic drift? In this study, we investigated patterns of genetic diversity and structure in the narrow endemic Hayden's ringlet butterfly. Specifically, we asked to what degree genetic structure in the Hayden's ringlet can be explained by isolation by distance, isolation by resistance (in the form of geographic or ecological barriers to migration between populations), and isolation by environment (in the form of differences in host plant availability and preference). We employed a genotyping-by-sequencing (GBS) approach coupled with host preference assays, Bayesian modeling, and population genomic analyses to answer these questions. Our results suggest that despite their restricted range, levels of genetic diversity in the Hayden's ringlet are comparable to those seen in more widespread butterfly species. Hayden's ringlets showed a strong preference for feeding on grasses relative to sedges, but neither larval preference nor potential host availability at sampling sites correlated with genetic structure. We conclude that geography, in the form of isolation by resistance and simple isolation by distance, was the major driver of contemporary patterns of differentiation in this narrow endemic species.

Keywords: *Coenonympha haydenii*, population structure, hierarchical Bayesian models, narrow endemism

Introduction

Determining the evolutionary processes underlying the development of population genetic structure can provide important insights into the causes and potential consequences of evolution. Patterns of genetic structure, or the organization of genetic diversity across geographic space, can help reveal contemporary gene flow and migratory routes (e.g., Gompert et al., 2021; Hemstrom et al., 2022), ecological specialization (e.g., Nosil et al., 2008; Ferrari et al., 2012; Chaturvedi et al., 2018; Michell et al., 2023), patterns of admixture (e.g., Prüfer et al., 2014), or even the initial stages of speciation (Mayr, 1942; Avise et al., 2000; Harvey et al., 2017). The development of genetic structure is driven by three major evolutionary processes: genetic drift, gene flow, and natural selection (Wright, 1931). But the degree to which each of these processes dominate—and what patterns of structure might arise as a result—depends heavily on geographic, ecological, and demographic conditions.

Narrow endemism (restriction of a species' range to a limited geographic area relative to dispersal capacity) is a condition that would, at first glance, appear to limit the potential for genetic structure to develop. Historically, it was predicted that narrow endemic species should show low levels of genetic diversity (Frankham, 1997; Soltis and Soltis, 1991). At small population sizes, genetic drift will more readily drive alleles to fixation, leading to loss of diversity over time (Wright, 1931; Gillespie, 2001; Montgomery et al., 2000; Rivera-Ortíz et al., 2015). Low levels of genetic diversity coupled with a narrowly limited geographic range (relative to dispersal capacity and habitat heterogeneticy) would seem to leave little genetic or geographic potential for differentiation to arise among populations. But a growing body of evidence suggests that endemic species—particularly plants—can show both high levels of genetic diversity (Forrest et al., 2017; Medrano and Herrera, 2008) as well as substantial genetic structure (Jiménez-Mejías et al., 2015; Hobbs et al., 2013; Turchetto et al., 2016). But is genetic structure in endemic species likely to be adaptive in nature (see Robitzch et al., 2023), or simply the result of limited gene flow and drift?

To help tease apart the processes driving the development of genetic structure, we can categorize the patterns of structure into three major cases. In the simplest case, population genetic structure can arise from a combination of geographic distance and genetic drift alone (Wright, 1943). This pattern is known as isolation by distance (IBD), and occurs when intrinsic limitations to dispersal lead to non-random mating and the accumulation of genetic differences across space via genetic drift, even in a perfectly uniform environment (Wright, 1943; Slatkin, 1993). If narrow endemic species experience a greater degree of genetic drift due population size limitations, then it might be predicted that IBD should more often be a key driver of patterns of structure in such species.

In cases where the environment connecting populations is not uniform, geographic or ecological barriers (e.g., mountains, rivers, low host availability) can reduce rates of gene flow among populations. Reduced rates of gene flow, in turn, can drive differentiation among isolated populations via genetic drift (Rivera-Ortíz et al., 2015). Geographically or ecologically favorable conditions, on the other hand, can create corridors of increased gene flow, homogenizing populations (Slatkin, 1987; Sharma et al., 2013). These conditions can result in patterns of genetic differentiation correlated with functional connectivity (i.e., heterogeneity in resistance of the landscape to gene flow) rather than physical distances (i.e. isolation by resistance, or IBR (McRae, 2006; Thomas et al., 2015; Moreno-Contreras et al., 2023). Because narrow endemics occur within a limited geographic range, in some cases there simply may not be enough environmental variation within a narrow endemic's range to result in substantial patterns of IBR. However, in many cases narrow endemic species are associated with ecologically unique environments and may be ecological specialists (for example, species endemic to white sands or serpentine soils; see Lavergne et al., 2004; Anacker et al., 2011; Metzler, 2014; Nery et al., 2023; Anacker, 2014). If narrow endemism is coupled with niche specialization, then narrow endemic species might be more likely to experience habitat fragmentation—especially if they also exhibit limited dispersal capacity—allowing isolation by resistance to develop even on a fine geographic scale.

Finally, if individual populations occupy ecologically divergent environments (as opposed to geographic or environmental barriers existing between two or more equivalent environments), natural selection can drive population divergence via local adaptation, resulting in a pattern known as isolation by environment (which subsumes isolation by adaptation) (Nosil et al., 2008, 2009; Orsini et al., 2013; Funk et al., 2011; Wang and Bradburd, 2014; Driscoe et al., 2019; Luna et al., 2023). Isolation by environment is specifically characterized genetic differentiation increasing with environmental differences between populations that are independent of geographic distances (Wang and Bradburd, 2014; Sexton et al., 2014). This can occur as a result of direct selection at loci affecting fitness, as well as indirect selection at neutral loci (Nosil et al., 2008, 2009). Thus, this pattern is the result of divergent selection coupled with reduced effective rates of gene flow (either via increased immigrant mortality or reduced hybrid fitness) increasing the potential for genetic hitchhiking, as well as limiting the extent to which gene flow erases the effects of natural selection (Nosil et al., 2008, 2009, 2008; Wang and Bradburd, 2014). Natural selection is more likely to overcome the effects of genetic drift when effective population sizes are large, and higher levels of standing genetic variation provide more raw material upon which natural selection can act. While narrow endemism implies that a species occurs over a limited geographic range, it does not imply that population sizes and genetic diversity levels are necessarily low. To the contrary, depending on various factors such as body size and local carrying capacity, small geographic ranges (from a human perspective) can support large, viable populations of some species. Moreover, the degree to which genetic diversity—particularly adaptive genetic diversity—decreases with range size reduction is predicted to be initially slow, with major reductions in diversity occurring only after the majority of a species' range has been eliminated (Exposito-Alonso et al., 2022). As narrow endemic species have increasingly been shown to harbor unexpectedly high levels of genetic diversity (Forrest et al., 2017; Medrano and Herrera, 2008), there is a possibility that patterns of genetic structure in such species could be adaptive in nature.

As narrow endemism is associated with increased extinction risk (Frankham, 1998; Pitman and Jørgensen, 2002), determining the degree to which genetic structure in narrow endemic species reflects patterns of local adaptation vs. genetic drift could have vital consequences for conservation. In species where genetic structure has arisen largely as a result of anthropogenic habitat fragmentation and drift (Ripperger et al., 2013; Johansson et al., 2007), populations might be better managed as a single unit. On the other hand, populations showing patterns of local adaptation might harbor vital adaptive variation, as well as show an increased risk for outbreeding depression due to high immigrant mortality or low hybrid fitness if not managed separately (Frankham et al., 2011). More broadly, understanding the nature of genetic structure in narrow endemic species could help shed light on the degree to which range size might influence the processes that dominate the development of genetic structure.

In this study, we characterize patterns of genetic diversity and structure in the narrow endemic Hayden's ringlet butterfly. The Hayden's ringlet, *Coenonympha haydenii*, is a brown Satyrid butterfly found only in mountain meadows and forest clearings of southwestern Montana, southeastern Idaho, and western Wyoming (i.e., the Greater Yellowstone Ecosystem) (Debinski and Pritchard, 2002; Pyle, 1981; Howe and Bauer, 1975; Scott, 1992). Known for both high local abundances (Caruthers and Debinski, 2006) and weak flying ability (Glassberg, 2001; Kaufman and Brock, 2003), it is possible that enough genetic variation and dispersal limitations could exist in this species to result in population genetic structure even at small spatial scales. Larvae of *C. haydenii* are thought to feed on one or more species of grasses (family Poaceae) or sedges (family Cyperaceae) (Debinski and Pritchard, 2002; Glassberg, 2001; Feltwell, 1993; Pyle, 1981). Female Hayden's ringlets are associated with moist, hydric meadows or bogs (Pyle, 1981; Scott, 1992), and population sizes decline during periods of drought (Debinski et al., 2013). This is consistent with the possibility that *C. haydenii* could be specialized on one or more endemic Yellowstone wetland species like sedges. Conversely, the congeneric and sympatric common ringlet (*Coenonympha tullia*) is

known to be a broad generalist, even feeding successfully on introduced species such as Kentucky bluegrass (Debinski and Pritchard, 2002). If the Hayden's ringlet is also able to utilize multiple host plants, it is possible natural selection could be driving local adaptation to host plant use across its range, particularly in disrupted environments where novel, invasive grass species dominate. These factors make the Hayden's ringlet an ideal system for investigating the processes driving patterns of genetic structure in narrow endemic species. Specifically, in this study we asked the following questions: (1) how much genetic diversity and structure exists within the Hayden's ringlet, and (2) to what degree is the development of genetic structure in this narrow endemic species associated with (a) geographic distance and genetic drift alone (isolation by distance), (b) geographic or ecological barriers to dispersal between populations (isolation by resistance), and/or (c) ecological differences and local adaptation to larval host plants among sites (isolation by environment). This will provide much-needed data regarding host use and population connectivity in an iconic Yellowstone butterfly, as well as contribute another example of how genetic structure can develop in a narrow endemic species.

Materials and Methods

Butterfly Sample Collection

Over the course of two years, we collected adult *C. haydenii* specimens of both sexes from 14 sampling sites across the species' range (see Fig. 1a). We surveyed for *C. haydenii* presence at two additional locations in the Yellowstone Plateau region (AVP and GLR, see Table S1) and along approximately 10 miles of trail on the John D. Rockefeller Jr. Memorial Parkway, but we only observed a single Hayden's ringlet across this entire region. Due to low abundance between Yellowstone National Park and Grand Teton National Park, we were unable to collect butterflies from this area. At each of the 14 sites where Hadyen's ringlet populations

were abundant, we collected an average of 27 butterflies per location (see Table 1 for specific sample sizes). Male butterflies were immediately frozen to preserve tissue for subsequent DNA extraction, while females were maintained temporarily in the lab for egg collection and oviposition preference assays and frozen afterwards. Butterfly specimens sampled within Yellowstone and Grand Teton National Parks were collected under permits YELL-2018-SCI-8064, YELL-2019-SCI-8064, GRTE-2018-SCI-0041, and GRTE-2019-SCI-0055.

DNA Sequencing, Alignment, and Variant Calling

We used Qiagen DNeasy 96 Blood and Tissue Kits to extract DNA from the thoracic tissue of 287 butterfly specimens representing 14 sampling locations (see Fig. 1 and Table 1). When available, an equal number of male and female specimens were chosen for sequencing from each site. Reduced-representation restriction-fragment based DNA libraries were prepared for genotyping-by-sequencing (GBS) following methods similar to those in Gompert et al. (2014a). Briefly, whole-genome DNA was digested using Mse1 and EcoR1 enzymes, ligated to custom barcode sequences, and amplified via PCR. Barcoded DNA fragments were then pooled across samples, purified, and size-selected using a BluePippin. DNA fragments between 300-450 bp were selected for sequencing. The resulting DNA fragment libraries were sequenced on the University of Texas Illumina HiSeq 4000 sequencing platform. The resulting DNA sequences were first filtered to remove PhiX sequences and poly-G tails. PhiX is a bacterial sequence introduced during HiSeq sequencing as an internal control. We used `SAMtools` version 1.10 and custom scripts to find and remove all reads that aligned to the PhiX reference genome, leaving 347,375,794 individual reads. Barcode sequences were then removed from these remaining reads using custom Perl scripts, allowing us to match each DNA sequence to the individual butterfly from which it came.

To date, no reference genome has been published for the Hayden's ringlet. In the absence of a full reference genome, we constructed a *de novo* set of reference contigs for *Coenonympha haydenii* using the program `CD-hit` version 4.8.1 (Li and Godzik, 2006). See

the Supplemental Information for further details regarding our construction of the reference contig set. Reads that were aligned to this reference contig set using `BWA` version 0.7.17-r1188 (Li and Durbin, 2009). We used the `BWA aln` algorithm, with the total number of mismatches allowed per read (-n) set to 5, or approximately 6% of each read. We set seed length (-l) equal to 20 bp, and the maximum allowed mismatches in the seed sequence (-k) equal to 2.

We identified sites with single nucleotide polymorphisms (SNPs) in our genomic data using `samtools` and `bcftools` version 1.9 (Li et al., 2009). We used the original consensus caller (-c) to call variants, and set the threshold probability (-p) for accepting variants to 0.01 (i.e., we only called variants if the posterior probability the nucleotide was invariant was less than 0.01). Variants were then filtered for quality using custom Perl scripts. We retained variable sites for which there were at least 2x more reads than the number of individuals we sequenced (i.e., mean coverage per $\geq$2x), contained a minimum of 10 reads for the alternative allele (to filter out possible sequencing errors), and had a phred-scaled mapping quality >30. We removed variant sites with base-quality rank-sum test, mapping-quality rank-sum test, and read-position rank-sum test p-values less than 0.001, 0.0001, and 0.001 respectively. We also removed any variable sites missing data for 20% or more of the individuals we sequenced. We set a maximum read depth of 8000 (3 standard deviations greater than the mean coverage level across loci) to remove possible paralogs/gene families, and removed all SNPs located less than 2 bps apart along a contig. After quality filtering, we were left with a total of 9313 SNPs for downstream analysis.

Assessing Patterns of Genetic Diversity and Structure

To measure overall levels of genetic diversity in the Hayden's ringlet, we calculated both Watterson's θ (θ_W) and nucleotide diversity (π). We estimated both diversity statistics and their 95% block bootstrap intervals using the program `ANGSD` version 0.933-71-g604e1a4 (Korneliussen et al., 2014), which uses the full set of aligned contigs (not our quality-filtered

SNP set) to account for uncertainty in the number of segregating sites present. We then calculated per-base-pair values of both θ_W and π based on the estimated number of bases sequenced from `ANGSD` using `R` version 4.2.2 (R Core Team, 2022).

To summarize patterns of genetic structure in the Hayden's ringlet, we first used the program `ENTROPY` version 2.0 to estimate admixture proportions (Gompert et al., 2014b; Shastry et al., 2021). `ENTROPY` is a program similar to the admixture model in `STRUCTURE`, but has the added feature of accounting for uncertainty in genotypes as captured by genotype likelihoods. It uses a Bayesian framework to co-estimate genotypes and the proportion of a particular individual's genome that would be derived from each of K hypothetical source populations. The purpose of this in our case was not to estimate the optimal value of K, but rather to assess patterns of coarse vs. fine-scale substructure within the species. To this end, we ran `ENTROPY` for all K-values between two and seven using our 9313-SNP set as input. For each value of K, we ran 10 Markov chain Monte Carlo (MCMC) chains with a 10,000-step burn-in period, 20,000 sampling iterations, a thinning interval of 5, and a Dirichlet initialization value of 50. As an additional summary of genetic structure, we then conducted a PCA in `R` using the (unscaled) posterior genotype estimates from `ENTROPY`.

We used Nei's F_{ST} (Nei, 1973) to quantify the magnitude of the genetic differentiation among the sampled populations. To calculate this, we first used `estpEM` version 0.1 (Soria-Carrasco et al., 2014) to obtain a maximum likelihood estimate of allele frequencies for each SNP (N = 9313) for each population (N = 14) of Hayden's ringlets we sampled. The program `estpEM` uses an expectation-maximization (EM) algorithm to account for uncertainty in genotypes arising from finite coverage and sequencing error (Soria-Carrasco et al., 2014). We set the tolerance level for EM convergence to 0.001, the maximum number of EM iterations to 20, and used our filtered genotype likelihood files split by population as input. With the allele frequency estimates for each population generated by `estpEM`, we then calculated pairwise Nei's F_{ST} values for each combination of populations, as well as overall F_{ST} across all populations. Briefly, we calculated the mean F_{ST} across all 9313 loci using the formula

$F_{ST} = \frac{1/L \sum_{i=1}^{L}(H_T - H_S)}{1/L \sum_{i=1}^{L}(H_T)}$ where H_T is the expected heterozygosity for the total population (i.e. across all subpopulation), H_S is the average of the expected heterozygosities within each subpopulation, and L is the number of loci (Lucek et al., 2019). These calculations were completed in `R`.

Tests for Isolation by Distance and Resistance

To determine the degree to which patterns of genetic structure in the Hayden's ringlet correlate to the geographic distances among sites (i.e. isolation by distance), we first conducted a Mantel test. We used the logit of pairwise F_{ST} and the natural log of euclidean distances among sites to produce our genetic and geographic distance matrices for comparison. The Mantel test was conducted in R version 4.2.2 using the package `vegan` version 2.6-4 (R Core Team, 2022; Oksanen et al., 2022). We used the Pearson correlation method, and ran the test for 999 permutations.

To identify geographic or ecological barriers to dispersal (i.e. isolation by resistance) separating *C. haydenii* sites, we used the statistical method Estimating Effective Migration Surfaces (`EEMS`), developed by Petkova et al. (2016). `EEMS` is based on the stepping-stone model of migration, and estimates effective migration rates by comparing the actual degree of genetic differentiation found among sites to the expectation under a null isolation-by-distance model. The model uses a resistance distance, which is a distance from circuit theory, to integrate over all possible dispersal paths between pairs of populations (Petkova et al., 2016). In contrast to some other circuit-theory based approaches (e.g., McRae, 2006; McRae et al., 2008), resistance distances are not defined *a priori* based on habitat features, but instead inferred from the data as part of model fitting (Petkova et al., 2016). This allows the identification of geographic regions among sites that might be serving as either environmental or geographic barriers or conduits to gene flow. We ran `EEMS` using a grid density of N-demes = 50, 100, and 150 demes. The number of demes corresponds to the

number of nodes EEMS produces in the triangular grid to which individual samples can be assigned. For each grid density level, we ran three MCMC chains of 4,000,000 steps each, a burn-in of 2,000,000 steps, and thinning interval of 9999.

Tests for Ecological Divergence Among Localities and Populations

We collected potential host plant specimens from 9 of our 14 sampling sites and species presence data from 12 of our 14 sampling sites as a measure of community assemblage. These data were collected to assess whether ecological differences among sampling sites correlate with genetic structure in the Hayden's ringlet (i.e. isolation by environment/adaptation). We additionally collected these data to serve as a record of potential host plants likely to be encountered by Hayden's ringlet populations across their range, and to inform which plant species would make the strongest candidates for oviposition and larval preference assays. The larvae of the Hayden's ringlet are suggested to feed generally on grasses (Debinski and Pritchard, 2002; Kaufman and Brock, 2003; Glassberg, 2001), at least some of which may overlap with the host genera used by the closely-related common ringlet butterfly, *C. tullia*. As such, we collected voucher specimens of each unique species of *Poa*, *Stipa*, and *Melica* grasses found in sampling site meadows where Hayden's ringlets were observed. All three of these grass genera are known to be suitable hosts for the congeneric and often sympatric common ringlet butterfly (*Coenonympha tullia*). It has also been suggested that Hayden's ringlets may be able to feed on sedges (family Cyperaceae) (Feltwell, 1993; Pyle, 1981), so we collected voucher specimens of all species of *Carex* sedges we found as well. After collection, plant specimens from different sites were classified by morphotype (or species where possible), and differences in community assemblage among sites were assessed using the Sørensen index. The Sørensen index measures the number of species shared between two sites as compared to the total number of species present across both sites, with greater weight given to shared than to non-shared species (Hao et al., 2019).

As a second test for potential host use differences among populations of Hayden's

ringlets, we conducted both female preference and larval preference assays and assessed differences in preference among populations. Since the preferred host(s) of the Hayden's ringlet are unknown, these assays were also conducted to determine whether this species is generally more likely to use grasses (Poaceae) or sedges (Cyperaceae) as their larval host. For these assays, we chose to compare preference for Hood's sedge (*Carex hoodii*) vs. Kentucky bluegrass (*Poa pratensis*). Both species are abundant throughout *C. haydenii's* range, and represent the two plant families Hayden's ringlets are hypothesized to feed upon: sedges (Cyperaceae) and grasses (Poaceae). Furthermore, the results of our plant community assemblage surveys showed that these species were the most well-represented members of their genus across our sampling sites, with *Carex hoodii* being observed at 10 out of 12 meadows and *Poa pratensis* being observed at 7 out of 12 meadows we collected Hayden's ringlets from.

Whereas feral Kentucky bluegrass was the most common species of *Poa* we found at our sampling sites for this study, it is also a non-native species. Kentucky bluegrass was introduced to the Great Plains and Rocky Mountain region as a forage crop for domestic livestock (McArthur et al., 1995). It is now one of the most abundant and widespread feral exotic plants in the region (Kay, 2001; Kauffman et al., 2023; McArthur et al., 1995), often reaching very high densities in meadows (Kay, 2001) and representing up to 40-50% of the vegetation cover in certain riparian regions across Yellowstone and the Grand Tetons (Kauffman et al., 2023). Feral Kentucky bluegrass is especially prevalent in meadows overgrazed by bison and elk (Kay, 2001; Kauffman et al., 2023), and now represents one of the dominant grass species in the Lamar Valley of Yellowstone National Park (Hunter et al., 2018).

Due to the high abundance of this exotic species throughout the range of the Hayden's ringlet and its propensity to alter the ecology and community structure of meadows it invades (Sanderson et al., 2017), the presence of Kentucky bluegrass could impose a strong selective pressure on Hayden's ringlet populations, setting the stage for local adaptation. The response of butterfly species to the presence of novel, exotic plant species is both well-documented and varied, with non-native species in some cases creating an ecological trap (for example,

if adult butterflies preferentially lay their eggs on an unsuitable, exotic host, e.g. Davis and Cipollini, 2014), and in other cases providing a lifeline for endangered species whose native host has gone extinct (Braga, 2023; Graves and Shapiro, 2003). In either case, the invasion of exotic plant species can have a substantial ecological and evolutionary impact on butterfly populations, even in remote areas. Indeed, in another Yellowstone area butterfly (genus *Lycaeides*), certain local populations have adapted to feed on feral roadside alfalfa (*Medicago sativa*), with alfalfa-adapted populations showing reduced oviposition preference for their native hosts (Forister et al., 2020; Chaturvedi et al., 2018). Similarly, the congener of the Hayden's ringlet, *Coenonympha tullia*, is both found in the greater Yellowstone area and known to successfully utilize Kentucky bluegrass as a larval host (Debinski and Pritchard, 2002)). Unlike the Hayden's ringlet, the common ringlet has rapidly expanded its range across the United States over the past 60 years (Wiernasz, 1983, 1989). It is possible the ability to feed on exotic species like Kentucky bluegrass could have played a role in this range expansion. Together, this makes Kentucky bluegrass both an ecologically relevant species to test as a potential host for the Hayden's ringlet, as well as a plant species with reasonable potential to be correlated with patterns of local adaptation in this species.

Finally, we chose harebell (*Campanula rotundifolia*) as a control group because it is a common herbaceous flower in the area (Craighead, 2005). Harebell is often found growing in meadows in association with grassland communities (Stevens et al., 2012), and thus may realistically be encountered by *C. haydenii* larvae in the wild. Harebell stem leaves are also long and narrow like those of grasses and sedges (Craighead, 2005; McGhan, 2023), which allowed us to control for leaf shape and size during our larval preference assays.

We conducted oviposition preference assays following standard procedures described in Forister et al. (2009), and assessed differences in preference across populations using a hierarchical Bayesian model. Briefly, we collected adult female butterflies from eight of our sampling sites (see Table 1) and placed them individually in plastic cups containing three plant samples each: Hood's sedge (*Carex hoodi*), Kentucky bluegrass (*Poa pratensis*), and

harebell (*Campanula rotundifolia*). All plant specimens used for these assays (from all three species) were collected from a meadow in greater Yellowstone area where Hayden's ringlets were abundant. Females were maintained in these cups for 72 hours, after which we counted the number of eggs adhered by each butterfly to each species of plant. Since female butterflies were given the choice of three plant species for oviposition, we modeled the number of eggs laid on each host plant multinomially. Specifically, we assumed the number of eggs laid on each host to follow the distribution multinomial$(P_{1:3}, n)$, where $P_{1:3}$ are the probabilities of oviposition on each host of the three host plants, and n is the total number of eggs laid. Each butterfly population was allowed its own oviposition probability values to account for potential differences in preference across populations. The oviposition probabilities ($P_{1:3}$) from each population were assigned a Dirichlet prior with $\alpha = \tau * S$. Here, the vector τ represents the global probability of oviposition on each host plant across all populations and S is a scaling factor that describes that variability in preference among populations. Finally, τ was assigned a Dirichlet hyperprior with $\alpha = 1$, and S a uniform hyperprior with lower and upper bounds of 1 and 200, respectively. We fit our model using `rjags` version 4.3.1 (Plummer, 2003, 2013). We ran three MCMC chains of 80,0000 sampling steps each, with a burn-in of 10,000 steps and thinning interval of 50. We checked convergence of the MCMC chains using the Gelman diagnostic (Gelman and Rubin, 1992).

After female oviposition preference assays were complete, all eggs laid in the oviposition cups were gently removed from their substrate and stored in vented petri dishes under ambient temperature and light conditions until they hatched (approximately 10 days). Within one day of hatching, we performed larval preference assays following standard protocols (Gómez Jiménez et al., 2014; Gamberale-Stille et al., 2014; Wang et al., 2017; Gu and Walter, 1999) to assess differences in larval feeding preferences across populations. We tested up to 40 neonate larvae each from 10 of our sampling sites (see Table 1). Larvae were placed in the center of petri dishes equidistant from three 1-cm long leaf segments representing each of our test species (Kentucky bluegrass, Hood's sedge, and harebell). We took pictures of

the leaf tissue flattened between glass slides both before and after the 72 hour herbivory trial with a Canon EOS M6 camera. We used the program `ImageJ` version 1.52A (Schneider et al., 2012) to trace outlines around each leaf image and calculate leaf surface area both before and after herbivory. The surface area lost by each leaf was calculated as the surface area before herbivory minus the surface area after herbivory (measured in cm^2). In addition, each leaf was manually assigned a binary value indicating whether signs of herbivory (i.e. jagged leaf margins) were observed (see the Supplemental Information for more details about our ImageJ protocol).

We estimated larval preferences among populations using a hierarchical Bayesian model. In our model, we assumed leaf area lost during the herbivory assays could be attributed to two main causes: (1) larval feeding, and (2) shrinkage of the leaf tissue due to moisture loss over time. We assumed total leaf area loss to follow a normal distribution with a mean and standard deviation as follows:

$$\text{total leaf area lost} \sim \text{N}(\text{shrinkage} + \text{herbivory} * \mathbb{1},\ \sigma_{loss}).$$

Here, $\mathbb{1}$ is a binary indicator set equal to 1 if herbivory was observed, and equal to 0 if no herbivory was observed. Thus, in cases where herbivory was observed, mean leaf area lost was defined as the sum of shrinkage plus larval herbivory. If no herbivory was observed, mean leaf area lost was defined as shrinkage only. We defined herbivory as following a normal distribution where the mean (μ_{herb}) and standard deviation (σ_{herb}) were allowed to vary by each unique plant species $\times$ butterfly population combination. Shrinkage was defined as following a normal distribution where the mean (μ_{shrink}) and standard deviation (σ_{shrink}) were allowed to vary by host plant only since the population each caterpillar was obtained from should have no effect on the amount of moisture lost by each leaf over time. The standard deviation parameters σ_{loss} , σ_{shrink}, and σ_{herb} were all assigned gamma priors with parameters $k = 2$ and $\theta = 0.1$, while μ_{shrink} was assigned a normal prior with $\mu = 0$ and

$\sigma = 2$. Meanwhile, μ_{herb} was defined as the sum of population and host effects multiplied by the probability of the caterpillar eating (P). The host effect was distributed normally with a mean of μ_{host} and standard deviation of 0.5, with μ_{host} assigned a normal prior of $\mathrm{N}(0, 20)$. A normal prior was also placed on the population effect, but with a sum-to-zero constraint for model indentifiability and a gamma prior for the standard deviation ($k = 2$, $\theta = 0.1$). For each host plant species $\times$ butterfly population combination, the total number of trials where larvae consumed leaf tissue was assigned to a binomial distribution with $n =$ number of trials and $p =$ the probability of a larva eating. We wrote this model in the language STAN (Stan Development Team, 2022b) and implemented it using the R-interface `RStan` version 2.21.5 (Stan Development Team, 2022a). We used a warm-up period of 15,000 steps and ran the model for 30,000 Hamiltonian Monte Carlo (HMC) steps.

Tests for Isolation by Environment

We quantified the degree to which patterns of genetic structure in the Hayden's ringlet are explained by geographic distance (i.e. isolation by distance) vs. ecological distance (i.e. ecological differences between the sites themselves, isolation by environment) using three different metrics of ecological distance: (i) the potential host plants available, (ii) oviposition preference distances, and (iii) larval herbivory preference distances. Each ecological distance was analyzed separately. For this, we used a Bayesian linear mixed model introduced by Gompert et al. (2014a), which extends a similar maximum-likelihood model from Clarke et al. (2002). This model accounts for the lack of independence among sampling site pairs (i.e. the genetic distance between populations A vs. B is not independent from the genetic distance between populations A vs. C because both comparisons include population A) (Gompert et al., 2014a). We modeled the effect of geographic and ecological distance on logit $\mathrm{F_{ST}}$ as follows:

$$\mathrm{logit}\left(\mathrm{F_{ST}}_{ij}\right) = \beta_0 + \beta_{geo} X_{ij}^{geo} + \beta_{eco} X_{ij}^{eco} + \lambda_i + \lambda_j.$$

Where X_{ij}^{geo} is the geographic distance (calculated as Euclidean distances) between each pair of sites and X_{ij}^{eco} is either (i) the potential host plant community dissimilarity (as measured by the Sørensen index), (ii) the median difference in oviposition preference for Kentucky bluegrass (*Poa pratensis*), or (iii) the median difference in larval preference for Kentucky bluegrass (*Poa pratensis*) for each pairwise combination of populations. Population random effects are represented by λ_i and λ_j. All distances were centered and standardized prior to running the model to account for differences in unit scale. We fit this model in R using `rjags` version 4.3.1 (Plummer, 2003, 2013). We ran 3 MCMC chains of 5000 sampling steps each, with a burn-in of 2000 steps and thinning interval of 5. We fit the full model shown above, along with sub-models including only geographic distance, only ecological distance, or neither distance (i.e. a null model). Deviance information criterion was used to compare the relative performance of the full model and sub-models for each ecological variable.

Results

Moderate genetic diversity and population structure exist in Hayden's Ringlet

Estimates of nucleotide diversity across populations of the Hayden's ringlet varied from $\pi = 0.00284$ at JSM (95% bootstrap interval 0.00281-0.00288) to $\pi = 0.00344$ at USL (95% bootstrap interval 0.00342-0.00347) (see Table 2). Estimates of θ_W were similar, ranging from a low of 0.00280 (95% bootstrap interval 0.00277-0.00282) at JSM to a high of 0.00360 (95% bootstrap interval 0.00359-0.00362) at BNP (see Table 2). Genetic structure across sites was moderate but notable, with an overall F_{ST} of 0.10. Pairwise F_{ST} comparisons (see Table 3) ranged from 0.0181 to 0.1191. The population pairs that showed the highest degree of genetic differentiation were USL vs. JSM ($F_{ST} = 0.1191$) and USL vs. PSP ($F_{ST} = 0.1071$). Meanwhile, the least-differentiated population pairs were TRL vs. BNP ($F_{ST} =$

437 0.0181) and HRP vs. MRF (F_{ST} = 0.0186). JSM and SKI, which are located very closely
438 in geographic space (∼5 km apart, see Fig. 1a and Table 3) nevertheless showed a degree of
439 differentiation comparable to population pairs much further apart in geographic space (F_{ST}
440 = 0.0609). Principal component analysis (PCA) shows individuals clustering by sampling
441 site (see Fig. 1b). In particular, we saw that PC 1 separates the northern Hayden's ringlet
442 populations from southern populations, while PC 2 separates the southern populations of
443 Hayden's ringlets along a NE to SW gradient. The PCA does not perfectly mirror the map
444 of our sampling locations, but is nevertheless suggestive of isolation by distance.

445 Admixture analysis (Fig. 2) showed the presence of meaningful structure across pop-
446 ulations of Hayden's ringlets across multiple levels of K. The most prominent pattern was
447 a clinal split between the northern and southern populations of Hayden's ringlets at K=2.
448 Higher values of K revealed additional substructure within the species. At $K = 3$, `ENTROPY`
449 split the southern populations of Hayden's ringlets along a North-South axis. In particu-
450 lar, we saw the southernmost population of Hayden's ringlets, PSP, being separated from
451 the remainder of the populations. Similarly, $K = 4$ split the northern populations across a
452 roughly West-East axis, separating northern populations east of the Gallatin mountain range
453 (BNP, HNV, TRL) from those west of this range (GNP, WTC). Higher levels of K continued
454 to refine the northeast-to-southwest clinal pattern seen across the southern populations of
455 Hayden's ringlets. A small number of individual butterflies (specifically from BCR, MRF,
456 GNP, and HNV) showed ancestry values that differed considerably from both the typical
457 values of their own population, as well as those of other populations we surveyed. This sug-
458 gests that these individuals could be migrants or of mixed ancestry. Overall, our admixture
459 analysis suggests that the greatest degree of genetic differentiation in the Hayden's ringlet
460 exists between northern and southern populations, with additional substructure occurring
461 within those geographic regions.

Isolation by distance and resistance both contribute to population structure in *C. haydenii*

We saw a strong signal for isolation by distance (see Fig. 4), with the Mantel test showing a significant and strong correlation between geographic and genetic distance in the Hayden's ringlet (R = 0.7, P = 0.001). In addition to isolation by distance (IBD), `EEMS` analysis showed several geographic areas with credibly increased or reduced relative migration rates (see Fig. 1c and 1d). Results for each of the three chains for grid sizes of 50, 100, and 150 were similar (see Fig. S1). There were several geographic areas within *C. haydenii*'s range where genetic differentiation among populations was either lower (low resistance) or higher (high resistance) than expected under a null IBD model alone, a pattern consistent with isolation by resistance. In particular, we saw a region of credibly reduced relative migration rates separating the northern and southern populations in our study, consistent with results from PC1 of the PCA (see Fig. 1d and 1b). This geographic region of credibly reduced gene flow produced by the `EEMS` model corresponds to the location of the Yellowstone plateau, roughly following the southern edge of the geothermally active Yellowstone volcanic area (see Fig. 1a). There was also a region of credibly increased relative migration connecting the majority of the southern populations of Hayden's ringlets with the exception of PSP, the southernmost population. This region of increased connectivity among southern Hayden's ringlet populations follows the river valley region known as Jackson hole, a low-elevation region between the Teton and Gros Ventre mountain regions (see Fig. 1a). The southernmost population (PSP), which showed credibly lower levels of gene flow with the remaining ringlet populations than expected under a null IBD model, is separated from the Jackson hole valley region by the Wyoming mountain range.

C. haydenii shows strong preference for grass host, but limited evidence for isolation by environment

All populations of *C. haydenii* we assessed laid credibly more eggs on Kentucky bluegrass than expected if females had no oviposition host preference (posterior probability [p.p.] for percent oviposition on *P. pratensis* > 33% > 0.98 for all, see Fig. 3). Oviposition rates on Kentucky bluegrass varied from a low of 51% (95% CI 33-65%) to a high of 74% (95% CI 61-87%) across populations. The median global preference for oviposition on Kentucky bluegrass across populations was 57% (95% CI 47-67%; p.p. percent oviposition on *P. pratensis* > 33% > 0.99), while the global preference for oviposition on Hood's sedge, *Carex hoodii*, was only 24% (95% CI 17-35%; p.p. percent oviposition on *C. hoodii* < 33% = 0.96). Median global preference for oviposition on harebell, our control species, was the lowest at only 17% (95% CI 10-26%; p.p. preference for *C. rotundifolia* < 0.33 > 0.99), 16 percentage points lower than expected if butterflies distributed their eggs equally across available substrates. The strength of oviposition preference varied credibly between several Hayden's ringlet population pairs, with both TRL and BTB showing credibly higher rates of oviposition on *Poa pratensis* than PIN, HRP, and WTC (p.p. > 0.99 for all six comparisons).

As with oviposition, Hadyen's ringlet larvae showed a strong preference for Kentucky bluegrass, *Poa pratensis*. The species-level preference for Kentucky bluegrass produced by the Bayesian model was 71% (95% CI 64%-79%), meaning we would expect 71% of the leaf tissue consumed by a randomly sampled group of Hayden's ringlet larvae to be from Kentucky bluegrass when given a choice of Kentucky bluegrass, Hood's sedge, and harebell. Unlike in the female oviposition assays, no harebell herbivory was observed from any of the larvae we assayed. All populations we assayed showed a trend toward consuming more grass (*Poa pratensis*) than sedge (*Carex hoodii*), with every population consuming credibly more grass than sedge (p.p. consumed more grass than sedge > 0.99) except SKI (p.p. SKI consumed more grass than sedge = 0.85).

The proportion of each host plant species eaten by larvae varied considerably by population. BCR and HRP showed the greatest preference for *Poa pratensis*, consuming 100% grass (95% CI 87-100% and 75-100% respectively) and 0% *Carex hoodii* sedge (95% CI 0-13% and 0-25% respectively). SKI, meanwhile, showed the lowest degree of herbivory preference, consuming 56.4% *Poa pratensis* grass (95% CI 44-71%) vs. 44% *Carex hoodii* sedge (95% CI 29-56%). Due to differences in total leaf tissue consumption across populations (25 out of 45 population pairs showed credible differences), we assessed differences in host preference across populations as differences in the proportion of grass vs. sedge leaf tissue consumed (see the Supplemental Information for details). We saw credible differences in the proportion of grass vs. sedge leaf tissue consumed for 21 out of 45 pairwise population comparisons. The pairs with the greatest differences in preference were BCR vs. SKI and HRP vs. TRL, with BCR and HRP consuming 42.1 (95% CI 27-55) and 41.9 (95% CI 17-55) percentage points more *Poa pratensis* grass and 42.1 (95% CI 27-55) and 41.9 (95% CI 17-55) percentage points less *Carex hoodii* than SKI and TRL respectively.

Despite finding credible differences in larval feeding and oviposition preferences across populations, we found no evidence that these differences correlated with genetic distances among Hayden's ringlet populations. The credible intervals for both the effect of larval preference and oviposition preference on logit(F_{ST}) overlapped zero (p.p. $\beta_{herb} > 0 = 22\%$; p.p. $\beta_{ovipos} > 0 = 36\%$, see Fig. 4c-d). This suggests that there is no measurable correlation between either larval host preference or oviposition preference for Kentucky bluegrass and genetic distances among Hayden's ringlet populations. The deviance information criterion (DIC) values for sub-models testing only the effect of larval preference (mean DIC = -21.65) or oviposition preference (mean DIC = -16) on genetic distance were substantially greater than for models and sub-models that included geographic distance as a variable (mean DIC ranged from -66 to -60). This suggests that our geographic distance models (both sub-models and the full models) better predict genetic distances in the Hayden's ringlet than models including oviposition or larval preference alone. Similarly, we found no measurable

effect of potential host community distance (as measured by the Sørensen index) on degree of genetic differentiation in the Hayden's ringlet (see Fig. 4b). The credible interval for β_{comm} overlapped zero (p.p. $\beta_{comm} > 0 = 73\%$), indicating there was no credible effect of the availability of *Poa*, *Stipa*, *Melica*, and *Carex* species across sites on genetic differentiation in the Hayden's ringlet. The DIC value for the sub-model including only host community as a variable was 15, while the sub-model and full model including geographic distance ranged from -63 to -62, again suggesting that the sub-model including only host community information was less predictive than models containing geographic distance information. Taken together, our data suggest that isolation by adaptation to the host plant communities we measured (a form of isolation by environment) is unlikely to be a driver of patterns of contemporary genetic structure in the Hayden's ringlet.

Discussion

In this study, we assessed patterns of genetic diversity and structure in the narrow endemic Hayden's ringlet. We also assessed patterns of oviposition and larval host preference, and used Bayesian methods and `EEMS` modeling to assess the role of isolation by distance, barriers to disperal (i.e. isolation by resistance), and potential host availability and preference (i.e. isolation by environment) contribute to population structure in this species. Our results indicate that despite range restriction, the Hayden's ringlet shows genetic diversity levels comparable to other more widely-distributed species. The Hayden's ringlet also appears to consistently prefer grass (*Poa pratensis*) over sedge (*Carex hoodii*), but this host association is unlikely to be driving patterns of population structure. Instead, we found that both isolation by distance and barriers to dispersal were most closely associated with genetic distances in this species. We discuss the implications of these results in more detail below.

Narrow endemism not associated with notable genetic diversity reduction in the Hayden's ringlet

Despite its restricted distribution, the Hayden's ringlet showed levels of genetic diversity comparable to more widely-distributed butterfly species. The average nucleotide diversity across Hayden's ringlet populations we sampled was $\pi = 0.003$, while nucleotide diversity in *Leptidea* sp., *Lycaeides melissa*, and *Parnassius mnemosyne* (all widely-distributed, non-migratory butterfly species) ranged from $\pi = 0.001$ to $\pi = 0.005$ (Talla et al., 2019; Gompert et al., 2014b; Talla et al., 2023). In contrast, both migratory monarchs (*Danaeus plexippus*) and non-migratory *Heliconius* sp. showed comparatively high nucleotide diversity ($\pi = 0.01$-0.06 and 0.020-0.28, respectively (Talla et al., 2020; Hemstrom et al., 2022; Martin et al., 2016; Kryvokhyzha, 2014). Migratory butterfly species have been shown to harbor higher levels of genetic diversity than non-migratory species in general, possibly due to greater population sizes and connectivity (García-Berro et al., 2023), so the substantial difference in nucleotide diversity between monarchs and Hayden's ringlets is not unexpected. However, *Heliconius* species are both non-migratory and have low dispersal ability (Kronforst and Fleming, 2001), so why this species group shows far higher genetic diversity levels than reported in other non-migratory species is unclear.

Many butterfly species have wide distributions, but are locally rare. The Hayden's ringlet, by contrast, is narrowly restricted in range, but locally prolific. Within their range, Hayden's ringlets are often so abundant they are the most common butterfly species surveyed (Caruthers and Debinski, 2006). High local abundances in the Hayden's ringlet could be one factor contributing to the maintenance of genetic diversity in this species. Conversely, poor dispersal (as seen in *Lycaeides melissa* and *Parnassius mnemosyne*) (Gompert et al., 2010; Talla et al., 2019; Gorbach and Kabanen, 2010) or poor connectivity among populations could lead to high levels of genetic drift, reducing nucleotide diversity estimates in more widespread butterfly species. In particular, the widely-distributed *Lycaeides melissa*

is known for low local population sizes, patchy distributions and metapopulation dynamics (Scott 1992; Gompert et al. 2010, 2012; but also see Guiney et al. 2010). While even low levels of gene flow can be enough to maintain nucleotide diversity across populations–even in the face of low effective population sizes for individual demes and substantial genetic drift (Whitlock and Barton, 1997; Gompert et al., 2021)–the more widespread a species is, the more likely it is that insurmountable geographic barriers to gene flow (even if this barrier is distance alone) might exist within their distribution. This could cause widespread species to behave more similarly to multiple, smaller demes with no gene flow amongst them than a single, panmictic population. Thus, the genetic diversity levels maintained in widely-distributed butterfly species might be expected to be more similar to those of geographically restricted species than global census sizes alone would suggest (Gompert et al., 2010). This could help explain why genetic diversity levels in non-migratory butterfly species do not appear to scale linearly with population size in nature (i.e. Lewontin's paradox) (Lewontin et al., 1974; Gompert et al., 2021; Charlesworth and Jensen, 2022).

In all, the similarity in diversity levels between the Hayden's ringlet vs. widely-distributed butterfly species suggest this is yet another case where narrow endemism is not associated with a notable reduction in genetic diversity. This adds to a growing body of research showing that even narrow endemic species can still harbor substantial genetic diversity (Forrest et al., 2017; Medrano and Herrera, 2008; Robitzch et al., 2023; Hobbs et al., 2013; Jiménez-Mejías et al., 2015). That said, nucleotide diversity amongst eukaryotes ranges from approximately $\pi = 0.001$ to $\pi = 0.15$ (Charlesworth and Jensen, 2022), placing the Hayden's ringlet firmly on the low end for eukaryotes as a whole. Other butterfly species with similar nucleotide diversity levels to the Hayden's ringlet have been targeted for conservation efforts (Talla et al., 2023). But neutral diversity should not be conflated with adaptive genetic diversity. Simulations suggest that loss of adaptive genetic diversity is likely to proceed more slowly than loss of neutral genetic diversity (Exposito-Alonso et al., 2022), so one must be cautious in presuming that species with low nucleotide diversity and

a limited distribution necessarily lack adaptive genetic potential. Nucleotide diversity levels alone are not sufficient to interpret whether or not the Hayden's ringlet is a species of conservation concern. While its narrow distribution put the Hayden's ringlet at greater risk of extirpation due to natural disasters (e.g., catastrophic fires or volcanic activity across the entire Yellowstone area), high local abundances coupled with genetic diversity levels comparable to more widely-distributed butterfly species suggests that the Hayden's ringlet is not necessarily at higher conservation risk due to genetic factors (i.e. inbreeding depression, etc.; see Frankham, 2005) than many other non-migratory, geographically widespread butterfly species.

Geography informs patterns of population genetic structure in the Hayden's ringlet

We saw clear evidence of population structure across the range of the Hayden's ringlet. The strongest signal of genetic differentiation was a geographic split between northern and southern populations of *C. haydenii*, with additional genetic substructure occurring within each of these groups.

The correlation between geographic and genetic distances in the Hayden's ringlet was R = 0.7, substantially higher than correlations seen in many other non-migratory butterfly species. Specifically, correlations between geographic and genetic distance for the Langue's metalmark (*Apodemia mormo langei*), heath fritillaries (*Melitaea athalia* and *Melitaea celadussa*), and checkerspots (*Euphydryas aurinia* and *Euphydryas editha*) ranged between R = 0.39 and R= 0.53 (Dupuis et al., 2018; Tahami et al., 2021; Mikheyev et al., 2013). This suggests that isolation by distance is able to explain a greater degree of the population structure observed in the Hayden's ringlet than in other non-migratory butterfly species. The high correlation between genetic and geographic distances in the Hayden's ringlet suggests much of the population structure observed in this narrow endemic species

can be attributed to genetic drift and limited dispersal.

Despite the clear patterns of genetic structure present in this species, F_{ST} values between populations of Hayden's ringlets were low to moderate. The scale of differentiation we observed is consistent with fine- to moderate-scale genetic population structure (F_{ST} between 0.01-0.2) seen in other non-migratory butterfly species (Talla et al., 2019; Pertoldi et al., 2021; Talla et al., 2023; Hinojosa et al., 2023), and on average greater than in migratory species like monarchs ($F_{ST} = 0.0001$) (Talla et al., 2020). While the F_{ST} values we observed may be considered low in other groups of organisms, in many cases F_{ST} values between nominal species of butterflies are not considerably greater than what we found within populations of the Hayden's ringlet (i.e. Talla et al., 2019; Tahami et al., 2021), and in some cases variation within butterfly species is higher than that observed between species. For example, in the El Segundo blue (*Euphilotes battoides allyni*), F_{ST} among populations of the same species ranged from 0.1 to 0.5 (Dupuis et al., 2020), while in heath fritillaries, F_{ST} between two nominal species (*Melitaea celadussa* and *Melitaea athalia*) was only 0.1-0.2 (Tahami et al., 2021). Thus, our results are clearly in-line with results from other butterfly species, and consistent with expectations for a non-migratory species with limited dispersal ability.

We saw several geographic regions with credibly increased or reduced relative migration rates in the Hayden's ringlet. The largest of these was a wide region of credibly reduced relative gene flow between northern and southern *C. haydenii* populations corresponding to the southern border of the Yellowstone plateau and John D. Rockefeller, Jr. Memorial Parkway. Despite having visited two additional sites (Avalanche Peak AVP, Grassy Lake Reservoir GLR; see Table S1 for coordinates) and surveyed approximately 10 miles of trail in this region, we found no viable populations of Hayden's ringlets connecting our northern and southern sampling sites. Much of the habitat in this region consisted of dense lodgepole pine monocultures and previous burn sites (Parmenter et al., 2003; Turner and Simard, 2017; Rothermel, 1994). Hayden's ringlets prefer open grassy meadows and sunny forest edges (Debinski and Pritchard, 2002; Kaufman and Brock, 2003), so this densely-forested region could

present an ecological barrier to migration. Regardless, the fact that our field observations are consistent with the results from our EEMS model suggests that this geographic region presents a true barrier to gene flow for the Hayden's ringlet, and that isolation by resistance contributes to patterns of genetic structure in this species. Interestingly, the geographic split we found between northern and southern *C. haydenii* populations corresponds to a similar boundary observed between northern *Lycaeides idas* populations and southern, admixed *Lycaeides* (Gompert et al., 2010, 2012). This suggests that a combination of geographic (elevation; mountain ranges) and ecological (forest type) conditions present in the John D. Rockefeller, Jr. Memorial Parkway may present a barrier to gene flow more generally, and could apply to other non-migratory butterfly species in the greater Yellowstone ecosystem as well.

Despite being a non-migratory species known for poor flight (Kaufman and Brock, 2003; Glassberg, 2001), we nevertheless saw evidence of long-distance dispersal in *C. haydenii*. Several individuals in our admixture analysis matched neither the population from which they were sampled, nor any other population we sampled. In particular, one individual each from MRF, GNP, and HNV in our admixture plots did not match the admixture proportions of any other butterflies we sampled. These individuals appear to be either of mixed origin or migrants from an area we did not sample. One individual from BCR, on the other hand, appears to be a migrant from PSP (or near PSP). The distance between PSP and BCR is over 65 km, indicating that long-distance dispersal does occur in *C. haydenii* at least occasionally. Hayden's ringlets are notoriously poor fliers (Glassberg, 2001; Kaufman and Brock, 2003), so we expect typical dispersal distances in the Hayden's ringlet to be similar to those reported for other poor dispersers like *Lycaeides melissa*, *Parnassius* sp., and *Heliconius* sp. (Gompert et al., 2010; Gorbach and Kabanen, 2010; Kronforst and Fleming, 2001), which rarely disperse further than 2 km during their lifetime. We suggest that the instances of long-distance dispersal we report here are likely a result of rare gene flow events such as butterflies being blown long distances during adverse weather conditions.

But as even small amounts of gene flow are sufficient to erase patterns of genetic differentiation, these occasional long-distance dispersal events likely still play a role in determining the magnitude of population genetic structure present in this species.

Strong preference for grass host, but no evidence of isolation by environment in the Hayden's ringlet

We observed strong oviposition and larval herbivory preference for Kentucky bluegrass (*Poa pratensis*) over Hood's sedge (*Carex hoodii*) in *C. haydenii*. Preference for grass was both strong and remarkably consistent, with all populations showing a credible preference for *Poa* in both oviposition and herbivory assays with the exception of SKI. While it has been previously suggested that Hayden's ringlets might feed on sedges due to their association with bogs and hydric habitats (Pyle, 1981; Scott, 1992), our evidence overwhelmingly points to grasses as being the preferred host of the Hayden's ringlet. However, the fact that larvae did often feed on both the sedge and grass host, while completely refusing the control host, suggests that Hayden's ringlets may accept more than one host, and are more likely generalist feeders like their congener the common ringlet, *C. tullia*, than narrow host-specialists (Scott, 1992; Debinski and Pritchard, 2002). This is consistent with preliminary host acceptance data we collected which showed that Hayden's ringlet larvae will consume tissue from many genera of grasses and sedges including *Stipa*, *Carex*, *Poa*, *Phleum*, and *Elymus* when given no other choice. Anecdotal evidence also suggests that Hayden's ringlet larvae can be reared to adulthood on *Carex* species (Stout, 2017), which would indeed suggest that the Hayden's ringlet is a broad generalist given their strong preference for *Poa*. That said, our study only compared only a single species of sedge with a single species of grass. It is possible these species alone are not sufficient to provide a full picture of *C. haydenii's* preference for grasses vs. sedges. Additional work is needed to further elucidate the degree of host specificity and preference in *C. haydenii*.

While the degree of preference for *Poa* varied credibly across populations, we saw no evidence of host-associated genetic differentiation across populations in the Hayden's ringlet. Neither potential host community differences nor differences in larval herbivory preference were predictive of genetic distances among Hayden's ringlet populations in our study. If the Hayden's ringlet is in fact a generalist feeder, and host use does not substantially impact larval fitness, then the composition and abundance of potential host species may have a limited effect on genetic differentiation. This could explain the absence of host-associated population structure we observed in this species. But how then do we interpret the phenotypic variation in host preference among populations we observed? It is possible the variation we saw reflects true variation for preference that exists among Hayden's ringlet populations in the wild. However, laboratory experiments must always be interpreted with caution with regard to their applicability in the field. In this case, we note that the Hayden's ringlet populations that showed the highest degree of herbivory preference also happened to be the populations that consumed the least total amount of plant material. Because our preference measure was scaled by total tissue consumed, the lower the total level of consumption, the more sensitive (and stochastic) our preference measure will be to small differences in herbivory. In other words, when total consumption is low, each bite of tissue a larva consumes will have a proportionally larger impact on preference than that same bite of tissue in a case where total consumption is high. Thus, in cases where total herbivory was low, herbivory preferences have the potential to appear exaggerated compared to cases where larvae ate a greater amount of total leaf tissue.

If the Hayden's ringlet is not limited to feeding on a narrow endemic Yellowstone-area plant species, what might be driving current patterns of range restriction in the Hayden's ringlet? Since we only assayed two species of potential hosts, one of which is an invasive species, we cannot definitively say that host specialization is not a driver of genetic differentiation and narrow endemism in the Hayden's ringlet. But preliminary work we conducted on larval performance showed that Hayden's ringlet larvae can survive on Kentucky

bluegrass through at least the 4th instar, at which point our larvae entered—and did not survive—diapause. Kentucky bluegrass, *Poa pratensis*, is one of the most widespread turf grass species in the United States (Huff et al., 2003). It is ubiquitous along roadsides and in lawns, occurs in all 50 states, and is highly invasive across the northern Great Plains and Yellowstone region, forming high-density feral populations throughout Yellowstone and the Grand Teton National Parks (DeKeyser et al., 2015; Hunter et al., 2018; McArthur et al., 1995; Kay, 2001; Kauffman et al., 2023). Counter to what might be expected if the presence of this novel, exotic species were exerting a strong selective pressure on Hayden's ringlet populations due to its unsuitability as a host, we saw no evidence that any of the populations we surveyed have developed a strong preference against feeding or ovipositing on this species. If Kentucky bluegrass is in fact a viable host for the Hayden's ringlet, it would strongly suggest that host specialization is not the key factor preventing contemporary range expansion in the Hayden's ringlet. Instead, other environmental factors not considered in this study, such as site elevation, temperature, rainfall, or forest cover, could play a greater role. In particular, the fact that populations of Hayden's ringlets decline during periods of drought (Debinski et al., 2013) suggest that the Hayden's ringlet might be restricted to wetter habitats. Perhaps a factor other than host plant use could be driving *C. haydenii*'s association with wetland areas. Hayden's ringlets overwinter as larvae, so it is possible moisture levels could have an effect on larval survival through winter diapause. On the other hand, it has also been suggested that the Hayden's ringlet could be a narrow endemic today simply because it is a remnant species left behind from a larger, pre-glaciation distribution, and it's range simply has not yet returned to its former size (Pyle, 1981). Unlike the Hayden's ringlet, the range of the common ringlet (*Coenonympa tullia*) is both able to use Kentucky bluegrass as a larval host and has expanded rapidly over the past 60 years (Debinski and Pritchard, 2002; Wiernasz, 1983, 1989). This expansion is thought to have been driven in part by a shift from univolitinism to multivoltinism (Wiernasz, 1983, 1989). Whether the Hayden's ringlet is univoltine or multivoltine does not appear to have been documented.

If the Hayden's ringlet is obligately univoltine, this could help explain why the Hayden's ringlet has remained endemic, while its congener has become widespread in distribution. More exploration of the life history and ecological requirements of the Hayden's ringlet are necessary to more fully understand the causes of genetic structure and narrow endemism in this species.

Conclusions

Despite their restricted range, we found that the Hayden's ringlet harbors genetic diversity levels comparable to geographically widespread, non-migratory butterfly species with similar dispersal ability. We found strong evidence that the Hayden's ringlet prefers grasses (*Poa*) over sedges (*Carex*) as a larval host, but work to determine the degree of host specificity in this species remains to be done. Geography, specifically isolation by distance and isolation by resistance (i.e. barriers to dispersal such as mountain ranges and/or regions of poor habitat) appear to be the driving factors producing patterns of population structure in the Hayden's ringlet. We found no evidence that either host preferences or host availability were correlated with genetic divergence, and it does not appear that isolation by environment is driving population divergence in this narrow endemic species. Instead, population structure in this species has likely developed largely via genetic drift, suggesting that the Hayden's ringlet would not necessarily benefit from being managed as more than one unit. That said, it is always possible that local adaptation to ecological factors we did not measure could be contributing to genetic structure in this species. Questions remain as to how evolutionary processes unfold in the face of narrow endemism, but in some cases at least, it appears that patterns of genetic diversity and structure in restricted vs. widespread species may not differ as greatly as one might initially suspect.

Acknowledgments

Thank you to Megan Brady for invaluable assistance with both field collections and preference experiments, to Angélica Traslaviña for many months of assistance with DNA extractions, and to both Camden Treat and Daniel Johnson for their extensive assistance cataloguing larval herbivory results. Support and resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged. This work was funded by the National Science Foundation (NSF GRFP awarded to AS, fellow 2017239847; NSF DEB 1844941 to ZG) and Utah State University.

Literature Cited

Anacker, B. L., 2014. The nature of serpentine endemism. American Journal of Botany 101:219–224.

Anacker, B. L., J. B. Whittall, E. E. Goldberg, and S. P. Harrison, 2011. Origins and consequences of serpentine endemism in the california flora. Evolution 65:365–376.

Avise, J. C. et al., 2000. Phylogeography: the History and Formation of Species. Harvard university press.

Braga, M. P., 2023. Are exotic host plants a life raft or a trap for butterflies? Current Opinion in Insect Science P. 101074.

Caruthers, J. C. and D. M. Debinski, 2006. Montane meadow butterfly species distributions in the greater yellowstone ecosystem. University of Wyoming National Park Service Research Center Annual Report 30:85–96.

Charlesworth, B. and J. D. Jensen, 2022. How can we resolve lewontin's paradox? Genome biology and evolution 14:evac096.

817 Chaturvedi, S., L. K. Lucas, C. C. Nice, J. A. Fordyce, M. L. Forister, and Z. Gompert,
818 2018. The predictability of genomic changes underlying a recent host shift in melissa blue
819 butterflies. Molecular Ecology 27:2651–2666.

820 Clarke, R. T., P. Rothery, and A. F. Raybould, 2002. Confidence limits for regression
821 relationships between distance matrices: estimating gene flow with distance. Journal of
822 Agricultural, Biological, and Environmental Statistics 7:361–372.

823 Craighead, C., 2005. Common Wildflowers of Grand Teton National Park. Grand Teton
824 Natural History Association.

825 Davis, S. L. and D. Cipollini, 2014. Do mothers always know best? oviposition mistakes
826 and resulting larval failure of pieris virginiensis on alliaria petiolata, a novel, toxic host.
827 Biological Invasions 16:1941–1950.

828 Debinski, D. M., J. C. Caruthers, D. Cook, J. Crowley, and H. Wickham, 2013. Gradient-
829 based habitat affinities predict species vulnerability to drought. Ecology 94:1036–1045.

830 Debinski, D. M. and J. Pritchard, 2002. A field guide to butterflies of the Greater Yellowstone
831 Ecosystem. Roberts Rinehart.

832 DeKeyser, E. S., L. A. Dennhardt, and J. Hendrickson, 2015. Kentucky bluegrass (poa
833 pratensis) invasion in the northern great plains: a story of rapid dominance in an endan-
834 gered ecosystem. Invasive Plant Science and Management 8:255–261.

835 Driscoe, A. L., C. C. Nice, R. W. Busbee, G. R. Hood, S. P. Egan, and J. R. Ott, 2019. Host
836 plant associations and geography interact to shape diversification in a specialist insect
837 herbivore. Molecular Ecology 28:4197–4211.

838 Dupuis, J. R., S. M. Geib, K. H. Osborne, and D. Rubinoff, 2020. Genomics confirms
839 surprising ecological divergence and isolation in an endangered butterfly. Biodiversity and
840 Conservation 29:1897–1921.

841 Dupuis, J. R., J. C. Oliver, B. M. Brunet, T. Longcore, J. J. Johnson, and F. A. Sperling,
842 2018. Genomic data indicate ubiquitous evolutionary distinctiveness among populations
843 of california metalmark butterflies. Conservation Genetics 19:1097–1108.

844 Exposito-Alonso, M., T. R. Booker, L. Czech, L. Gillespie, S. Hateley, C. C. Kyriazis, P. L.
845 Lang, L. Leventhal, D. Nogues-Bravo, V. Pagowski, et al., 2022. Genetic diversity loss in
846 the anthropocene. Science 377:1431–1435.

847 Feltwell, J., 1993. The Encyclopedia of Butterflies. Prentice Hall.

848 Ferrari, J., J. A. West, S. Via, and H. C. J. Godfray, 2012. Population genetic structure and
849 secondary symbionts in host-associated populations of the pea aphid complex. Evolution
850 66:375–390.

851 Forister, M. L., C. C. Nice, J. A. Fordyce, and Z. Gompert, 2009. Host range evolution
852 is not driven by the optimization of larval performance: the case of lycaeides melissa
853 (lepidoptera: Lycaenidae) and the colonization of alfalfa. Oecologia 160:551–561.

854 Forister, M. L., S. A. Yoon, C. S. Philbin, C. D. Dodson, B. Hart, J. G. Harrison, O. Shelef,
855 J. A. Fordyce, Z. H. Marion, C. C. Nice, et al., 2020. Caterpillars on a phytochemical
856 landscape: The case of alfalfa and the melissa blue butterfly. Ecology and Evolution
857 10:4362–4374.

858 Forrest, A., M. Escudero, M. Heuertz, Y. Wilson, E. Cano, and P. Vargas, 2017. Testing the
859 hypothesis of low genetic diversity and population structure in narrow endemic species:
860 the endangered antirrhinum charidemi (plantaginaceae). Botanical Journal of the Linnean
861 Society 183:260–270.

862 Frankham, R., 1997. Do island populations have less genetic variation than mainland pop-
863 ulations? Heredity 78:311–327.

———, 1998. Inbreeding and extinction: island populations. Conservation biology 12:665–675.

———, 2005. Genetics and extinction. Biological conservation 126:131–140.

Frankham, R., J. D. Ballou, M. D. Eldridge, R. C. Lacy, K. Ralls, M. R. Dudash, and C. B. Fenster, 2011. Predicting the probability of outbreeding depression. Conservation Biology 25:465–475.

Funk, D. J., S. P. Egan, and P. Nosil, 2011. Isolation by adaptation in neochlamisus leaf beetles: Host-related selection promotes neutral genomic divergence. Molecular Ecology 20:4671–4682.

Gamberale-Stille, G., L. Söderlind, N. Janz, and S. Nylin, 2014. Host plant choice in the comma butterfly–larval choosiness may ameliorate effects of indiscriminate oviposition. Insect science 21:499–506.

García-Berro, A., V. Talla, R. Vila, H. K. Wai, D. Shipilina, K. G. Chan, N. E. Pierce, N. Backström, and G. Talavera, 2023. Migratory behaviour is positively associated with genetic diversity in butterflies. Molecular Ecology 32:560–574.

Gelman, A. and D. B. Rubin, 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7:457–472.

Gillespie, J. H., 2001. Is the population size of a species relevant to its evolution? Evolution 55:2161–2169.

Glassberg, J., 2001. Butterflies through binoculars: the west. Oxford University Press.

Gómez Jiménez, M. I., C. E. Sarmiento, M. F. Díaz, A. Chautá, A. Peraza, A. Ramírez, and K. Poveda, 2014. Oviposition, larval preference, and larval performance in two polyphagous species: does the larva know best? Entomologia Experimentalis et Applicata 153:24–33.

888 Gompert, Z., L. K. Lucas, C. A. Buerkle, M. L. Forister, J. A. Fordyce, and C. C. Nice,
889 2014a. Admixture and the organization of genetic diversity in a butterfly species complex
890 revealed through common and rare genetic variants. Molecular ecology 23:4555–4573.

891 ———, 2014b. Admixture and the organization of genetic diversity in a butterfly species
892 complex revealed through common and rare genetic variants. Molecular ecology 23:4555–
893 4573.

894 Gompert, Z., L. K. Lucas, J. A. Fordyce, M. L. Forister, and C. C. Nice, 2010. Secondary
895 contact between lycaeides idas and l. melissa in the rocky mountains: extensive admixture
896 and a patchy hybrid zone. Molecular ecology 19:3171–3192.

897 Gompert, Z., L. K. Lucas, C. C. Nice, J. A. Fordyce, M. L. Forister, and C. A. Buerkle, 2012.
898 Genomic regions with a history of divergent selection affect fitness of hybrids between two
899 butterfly species. Evolution 66:2167–2181.

900 Gompert, Z., A. Springer, M. Brady, S. Chaturvedi, and L. K. Lucas, 2021. Genomic time-
901 series data show that gene flow maintains high genetic diversity despite substantial genetic
902 drift in a butterfly species. Molecular Ecology 30:4991–5008.

903 Gorbach, V. and D. Kabanen, 2010. Spatial organization of the clouded apollo population
904 (parnassius mnemosyne) in onega lake basin. Entomological Review 90:11–22.

905 Graves, S. D. and A. M. Shapiro, 2003. Exotics as host plants of the california butterfly
906 fauna. Biological conservation 110:413–433.

907 Gu, H. and G. Walter, 1999. Is the common sowthistle (sonchus oleraceus) a primary host
908 plant of the cotton bollworm, helicoverpa armigera (lep., noctuidae)? oviposition and
909 larval performance. Journal of Applied Entomology 123:99–105.

910 Guiney, M. S., D. A. Andow, and T. T. Wilder, 2010. Metapopulation structure and dy-
911 namics of an endangered butterfly. Basic and Applied Ecology 11:354–362.

912 Hao, M., J. J. Corral-Rivas, M. S. González-Elizondo, K. N. Ganeshaiah, M. G. Nava-
913 Miranda, C. Zhang, X. Zhao, and K. Von Gadow, 2019. Assessing biological dissimilarities
914 between five forest communities. Forest Ecosystems 6:1–8.

915 Harvey, M. G., G. F. Seeholzer, B. T. Smith, D. L. Rabosky, A. M. Cuervo, and R. T. Brum-
916 field, 2017. Positive association between population genetic differentiation and speciation
917 rates in new world birds. Proceedings of the National Academy of Sciences 114:6328–6333.

918 Hemstrom, W. B., M. G. Freedman, M. P. Zalucki, S. R. Ramírez, and M. R. Miller,
919 2022. Population genetics of a recent range expansion and subsequent loss of migration in
920 monarch butterflies. Molecular Ecology 31:4544–4557.

921 Hinojosa, J. C., C. Montiel-Pantoja, M. Sanjurjo-Franch, I. Martínez-Pérez, K. M. Lee,
922 M. Mutanen, and R. Vila, 2023. Diversification linked to larval host plant in the butterfly
923 eumedonia eumedon. Molecular Ecology .

924 Hobbs, J.-P. A., L. Van Herwerden, D. R. Jerry, G. P. Jones, and P. L. Munday, 2013. High
925 genetic diversity in geographically remote populations of endemic and widespread coral
926 reef angelfishes (genus: Centropyge). Diversity 5:39–50.

927 Howe, W. H. and D. L. Bauer, 1975. the Butterflies of north America. Garden City, NY:
928 Doubleday.

929 Huff, D. R., M. Casler, and R. Duncan, 2003. Kentucky bluegrass. Turfgrass biology,
930 genetics, and breeding. Wiley, Hoboken, NJ Pp. 27–38.

931 Hunter, H. E., P. O. Husby, J. Fidel, and J. C. Mosley, 2018. Ecological health of grasslands
932 and sagebrush steppe on the northern yellowstone range. Rangelands 40:212–223.

933 Jiménez-Mejías, P., M. Fernández-Mazuecos, M. E. Amat, and P. Vargas, 2015. Narrow
934 endemics in european mountains: high genetic diversity within the monospecific genus

935 pseudomisopates (plantaginaceae) despite isolation since the late pleistocene. Journal of
936 Biogeography 42:1455–1468.

937 Johansson, M., C. R. Primmer, and J. Merilä, 2007. Does habitat fragmentation reduce
938 fitness and adaptability? a case study of the common frog (rana temporaria). Molecular
939 Ecology 16:2693–2700.

940 Kauffman, J. B., D. L. Cummings, C. Kauffman, R. L. Beschta, J. Brooks, K. MacNeill,
941 and W. J. Ripple, 2023. Bison influences on composition and diversity of riparian plant
942 communities in yellowstone national park. Ecosphere 14:e4406.

943 Kaufman, K. and J. Brock, 2003. Field guide to butterflies of north america. Hillstar Editions
944 LC .

945 Kay, C., 2001. Long-term aspen exclosures in the Yellowstone ecosystem. US Department
946 of Agriculture, Forest Service, Rocky Mountain Research ....

947 Korneliussen, T. S., A. Albrechtsen, and R. Nielsen, 2014. ANGSD: Analysis of next gener-
948 ation sequencing data. BMC Bioinformatics 15:356.

949 Kronforst, M. R. and T. H. Fleming, 2001. Lack of genetic differentiation among widely
950 spaced subpopulations of a butterfly with home range behaviour. Heredity 86:243–250.

951 Kryvokhyzha, D., 2014. Whole genome resequencing of heliconius butterflies revolutionizes
952 our view of the level of admixture between species.

953 Lavergne, S., J. D. Thompson, E. Garnier, and M. Debussche, 2004. The biology and ecology
954 of narrow endemic and widespread plants: a comparative study of trait variation in 20
955 congeneric pairs. Oikos 107:505–518.

956 Lewontin, R. C. et al., 1974. The genetic basis of evolutionary change, vol. 560. Columbia
957 University Press New York.

Li, H. and R. Durbin, 2009. Fast and accurate short read alignment with burrows–wheeler transform. bioinformatics 25:1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin, 2009. The sequence alignment/map format and samtools. Bioinformatics 25:2078–2079.

Li, W. and A. Godzik, 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659.

Lucek, K., Z. Gompert, and P. Nosil, 2019. The role of structural genomic variants in population differentiation and ecotype formation in timema cristinae walking sticks. Molecular ecology 28:1224–1237.

Luna, L. W., L. N. Naka, G. Thom, L. L. Knowles, A. O. Sawakuchi, A. Aleixo, and C. C. Ribas, 2023. Late pleistocene landscape changes and habitat specialization as promoters of population genomic divergence in Amazonian floodplain birds. Molecular Ecology 32:214–228.

Martin, S. H., M. Möst, W. J. Palmer, C. Salazar, W. O. McMillan, F. M. Jiggins, and C. D. Jiggins, 2016. Natural selection and genetic diversity in the butterfly heliconius melpomene. Genetics 203:525–541.

Mayr, E., 1942. Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press.

McArthur, E. D., A. C. Blauer, S. B. Monsen, and S. C. Sanderson, 1995. Plant inventory, succession, and reclamation alternatives on disturbed lands in grand teton national park. *in* Proceedings: Wildland Shrub and Arid Land Restoration Symposium, Pp. 343–358. Intermountain Research Station.

McGhan, P. J. R., 2023. United states department of agriculture forest service, plant of the week: Harebell (campanula rotundifolia l.). https://www.fs.usda.gov/wildflowers/plant-of-the-week/campanula_rotundifolia.shtml. Accessed on 2023-01-27.

McRae, B. H., 2006. Isolation by resistance. Evolution 60:1551–1561.

McRae, B. H., B. G. Dickson, T. H. Keitt, and V. B. Shah, 2008. Using circuit theory to model connectivity in ecology, evolution, and conservation. Ecology 89:2712–2724.

Medrano, M. and C. M. Herrera, 2008. Geographical structuring of genetic diversity across the whole distribution range of narcissus longispathus, a habitat-specialist, mediterranean narrow endemic. Annals of Botany 102:183–194.

Metzler, E. H., 2014. The remarkable endemism of moths at white sands national monument in new mexico, usa, with special emphasis on gelechioidea (lepidoptera). Journal of Asia-Pacific Biodiversity 7:e1–e5.

Michell, C. T., N. Wagner, M. Mutanen, K. M. Lee, and T. Nyman, 2023. Genomic evidence for contrasting patterns of host-associated genetic differentiation across shared host-plant species in leaf-and bud-galling sawflies. Molecular Ecology .

Mikheyev, A. S., C. S. McBride, U. G. Mueller, C. Parmesan, M. R. Smee, C. Stefanescu, B. Wee, and M. C. Singer, 2013. Host-associated genomic differentiation in congeneric butterflies: Now you see it, now you do not. Molecular ecology 22:4753–4766.

Montgomery, M. E., L. M. Woodworth, R. K. Nurthen, D. M. Gilligan, D. A. Briscoe, and R. Frankham, 2000. Relationships between population size and loss of genetic diversity: comparisons of experimental results with theoretical predictions. Conservation Genetics 1:33–43.

Moreno-Contreras, I., A. Llanes-Quevedo, L. A. Sánchez-González, M. d. C. Arizmendi, and

A. G. Navarro-Sigüenza, 2023. Isolation by resistance explains genetic diversity in the Arremon brushfinches of northern Mesoamerica. Molecular Ecology 32:3450–3470.

Nei, M., 1973. Analysis of gene diversity in subdivided populations. Proceedings of the national academy of sciences 70:3321–3323.

Nery, E. K., M. K. Caddah, M. F. Santos, and A. Nogueira, 2023. The evolution of ecological specialization underlies plant endemism in the atlantic forest. Annals of Botany P. mcad029.

Nosil, P., S. P. Egan, and D. J. Funk, 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. Evolution 62:316–336.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos, 2009. Divergent selection and heterogeneous genomic divergence. Molecular ecology 18:375–402.

Oksanen, J., G. Simpson, F. Blanchet, R. Kindt, P. Legendre, P. Minchin, R. O'hara, P. Solymos, M. Stevens, E. Szoecs, et al., 2022. Vegan: Community ecology package, r package version 2.6-4.

Orsini, L., J. Vanoverbeke, I. Swillen, J. Mergeay, and L. De Meester, 2013. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Molecular ecology 22:5983–5999.

Parmenter, A. W., A. Hansen, R. E. Kennedy, W. Cohen, U. Langner, R. Lawrence, B. Maxwell, A. Gallant, and R. Aspinall, 2003. Land use and land cover change in the greater yellowstone ecosystem: 1975–1995. Ecological Applications 13:687–703.

Pertoldi, C., A. Ruiz-Gonzalez, S. Bahrndorff, N. Renee Lauridsen, T. Nisbeth Henriksen, A. Eskildsen, and T. T. Høye, 2021. Strong isolation by distance among local populations

of an endangered butterfly species (euphydryas aurinia). Ecology and Evolution 11:12790–12800.

Petkova, D., J. Novembre, and M. Stephens, 2016. Visualizing spatial population structure with estimated effective migration surfaces. Nature genetics 48:94–100.

Pitman, N. C. and P. M. Jørgensen, 2002. Estimating the size of the world's threatened flora. Science 298:989–989.

Plummer, M., 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *in* Proceedings of the 3rd International Workshop on Distributed Statistical Computing, vol. 124, P. 10. Vienna, Austria.

———, 2013. rjags: Bayesian graphical models using MCMC. R package version 3.

Prüfer, K., F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. De Filippo, et al., 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505:43–49.

Pyle, R. M., 1981. The Audubon society field guide to North American butterflies. Knopf; distributed by Random House.

R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org`.

Ripperger, S. P., M. Tschapka, E. K. Kalko, B. Rodriguez-Herrera, and F. Mayer, 2013. Life in a mosaic landscape: anthropogenic habitat fragmentation affects genetic population structure in a frugivorous bat species. Conservation Genetics 14:925–934.

Rivera-Ortíz, F., R. Aguilar, M. Arizmendi, M. Quesada, and K. Oyama, 2015. Habitat fragmentation and genetic variability of tetrapod populations. Animal Conservation 18:249–258.

1050 Robitzch, V., P. Saenz-Agudelo, T. J. Alpermann, B. Frédérich, and M. L. Berumen, 2023.
1051 Contrasting genetic diversity and structure between endemic and widespread damselfishes
1052 are related to differing adaptive strategies. Journal of Biogeography 50:380–392.

1053 Rothermel, R. C., 1994. Fire growth maps for the 1988 Greater Yellowstone Area fires, vol.
1054 304. US Department of Agriculture, Forest Service, Intermountain Research Station.

1055 Sanderson, M. A., H. Johnson, M. A. Liebig, J. R. Hendrickson, and S. E. Duke, 2017.
1056 Kentucky bluegrass invasion alters soil carbon and vegetation structure on northern mixed-
1057 grass prairie of the united states. Invasive Plant Science and Management 10:9–16.

1058 Schneider, C. A., W. S. Rasband, and K. W. Eliceiri, 2012. Nih image to imagej: 25 years
1059 of image analysis. Nature methods 9:671–675.

1060 Scott, J. A., 1992. The butterflies of North America: a natural history and field guide.
1061 Stanford University Press.

1062 Sexton, J. P., S. B. Hangartner, and A. A. Hoffmann, 2014. Genetic isolation by environment
1063 or distance: which pattern of gene flow is most common? Evolution 68:1–15.

1064 Sharma, S., T. Dutta, J. E. Maldonado, T. C. Wood, H. S. Panwar, and J. Seiden-
1065 sticker, 2013. Forest corridors maintain historical gene flow in a tiger metapopulation
1066 in the highlands of central india. Proceedings of the Royal Society B: Biological Sciences
1067 280:20131506.

1068 Shastry, V., P. E. Adams, D. Lindtke, E. G. Mandeville, T. L. Parchman, Z. Gompert, and
1069 C. A. Buerkle, 2021. Model-based genotype and ancestry estimation for potential hybrids
1070 with mixed-ploidy. Molecular ecology resources 21:1434–1451.

1071 Slatkin, M., 1987. Gene flow and the geographic structure of natural populations. Science
1072 236:787–792.

———, 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279.

Soltis, P. S. and D. E. Soltis, 1991. Genetic variation in endemic and widespread plant species. Aliso: A Journal of Systematic and Floristic Botany 13:215–223.

Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, et al., 2014. Stick insect genomes reveal natural selection's role in parallel speciation. Science 344:738–742.

Stan Development Team, 2022a. RStan: the R interface to Stan. URL `https://mc-stan.org/`. R package version 2.21.5.

———, 2022b. Stan modeling language users guide and reference manual. URL `https://mc-stan.org/`. Version 2.30.

Stevens, C. J., J. Wilson, and H. A. McAllister, 2012. Biological flora of the british isles: Campanula rotundifolia. Journal of Ecology 100:821–839.

Stout, T., 2017. Personal communication. `http://www.raisingbutterflies.org/raising_butterflies_301/`.

Tahami, M. S., V. Dincă, K. M. Lee, R. Vila, M. Joshi, M. Heikkilä, L. Dapporto, S. Schmid, P. Huemer, and M. Mutanen, 2021. Genomics reveal admixture and unexpected patterns of diversity in a parapatric pair of butterflies. Genes 12:2009.

Talla, V., A. Johansson, V. Dincă, R. Vila, M. Friberg, C. Wiklund, and N. Backström, 2019. Lack of gene flow: narrow and dispersed differentiation islands in a triplet of leptidea butterfly species. Molecular ecology 28:3756–3770.

Talla, V., V. Mrazek, J. Höglund, and N. Backström, 2023. Whole genome re-sequencing uncovers significant population structure and low genetic diversity in the endangered clouded apollo (parnasssius mnemosyne) in sweden. Conservation Genetics Pp. 1–10.

1097 Talla, V., A. A. Pierce, K. L. Adams, T. J. de Man, S. Nallu, F. X. Villablanca, M. R.
1098 Kronforst, and J. C. de Roode, 2020. Genomic evidence for gene flow between monarchs
1099 with divergent migratory phenotypes and flight performance. Molecular ecology 29:2567–
1100 2582.

1101 Thomas, L., W. J. Kennington, M. Stat, S. P. Wilkinson, J. T. Kool, and G. A. Kendrick,
1102 2015. Isolation by resistance across a complex coral reef seascape. Proceedings of the
1103 Royal Society B: Biological Sciences 282:20151217.

1104 Turchetto, C., A. L. A. Segatto, G. Mäder, D. M. Rodrigues, S. L. Bonatto, and L. B.
1105 Freitas, 2016. High levels of genetic diversity and population structure in an endemic and
1106 rare species: implications for conservation. AoB Plants 8.

1107 Turner, M. G. and M. Simard, 2017. Using spatial statistics and landscape metrics to com-
1108 pare disturbance mosaics. Learning Landscape Ecology: A Practical Guide to Concepts
1109 and Techniques Pp. 175–190.

1110 Wang, I. J. and G. S. Bradburd, 2014. Isolation by environment. Molecular ecology 23:5649–
1111 5662.

1112 Wang, Y., Y. Ma, D.-S. Zhou, S.-X. Gao, X.-C. Zhao, Q.-B. Tang, C.-Z. Wang, and J. J.
1113 van Loon, 2017. Higher plasticity in feeding preference of a generalist than a specialist:
1114 experiments with two closely related helicoverpa species. Scientific Reports 7:1–12.

1115 Whitlock, M. C. and N. H. Barton, 1997. The effective size of a subdivided population.
1116 Genetics 146:427–441.

1117 Wiernasz, D. C., 1983. Range expansion and rapid evolution in Coenonympha tullia (Lepi-
1118 doptera): Ecological and genetic change in a new environment. Princeton University.

1119 ———, 1989. Ecological and genetic correlates of range expansion in coenonympha tullia.
1120 Biological Journal of the Linnean Society 38:197–214.

Wright, S., 1931. Evolution in Mendelian populations. Genetics 16:97.

———, 1943. Isolation by distance. Genetics 28:114.

Data Accessibility and Benefit-Sharing

Data Accessibility Statement

Raw sequence reads have been deposited in NCBI's SRA (BioProject PRJNA1036281). Scripts, ecological data and downstream genetic data are available from Dryad (`https://doi.org/10.5061/dryad.zw3r228g3`)

Benefit-Sharing Statement

Benefits from this research accrue from the sharing of our data and results on public databases as described above.

Author Contributions

AS and ZG designed the research. AS performed the research. AS and ZG analyzed the data. AS wrote the paper, with guidance and editing from ZG.

Tables and Figures

Table 1: Collection locations and sample sizes for the total number of adult butterflies collected from each site, the total number of specimens from which DNA was extracted and sequenced, the number of female butterflies for which oviposition preference assays were conducted, the number of female butterflies that produced offspring for the larval herbivory assays, and the total number of larvae for which herbivory assays were conducted.

	Latitude	Longitude	Butterflies collected	DNA sequenced	Oviposition pref.	Mothers of larvae	Larval pref.
BCR	43.3007	-110.5530	30	24	0	6	28
BNP	44.9337	-110.7212	30	24	0	0	0
BTB	43.6382	-110.6820	10	10	9	7	40
GNP	45.4323	-111.2245	35	24	2	2	5
HNV	44.6823	-110.4945	30	23	0	0	0
HRP	43.8957	-110.6427	26	25	9	5	21
JSM	43.5107	-110.9862	5	5	0	0	0
MRF	43.8547	-110.3918	36	24	0	7	40
PIN	43.7398	-109.9762	33	24	12	7	40
PSP	42.7483	-110.8398	26	23	0	0	0
SKI	43.5094	-110.9227	48	24	12	3	16
TRL	44.9019	-110.1291	30	24	13	7	40
USL	43.5829	-110.3328	9	9	2	1	11
WTC	44.7849	-111.3088	31	24	7	6	40

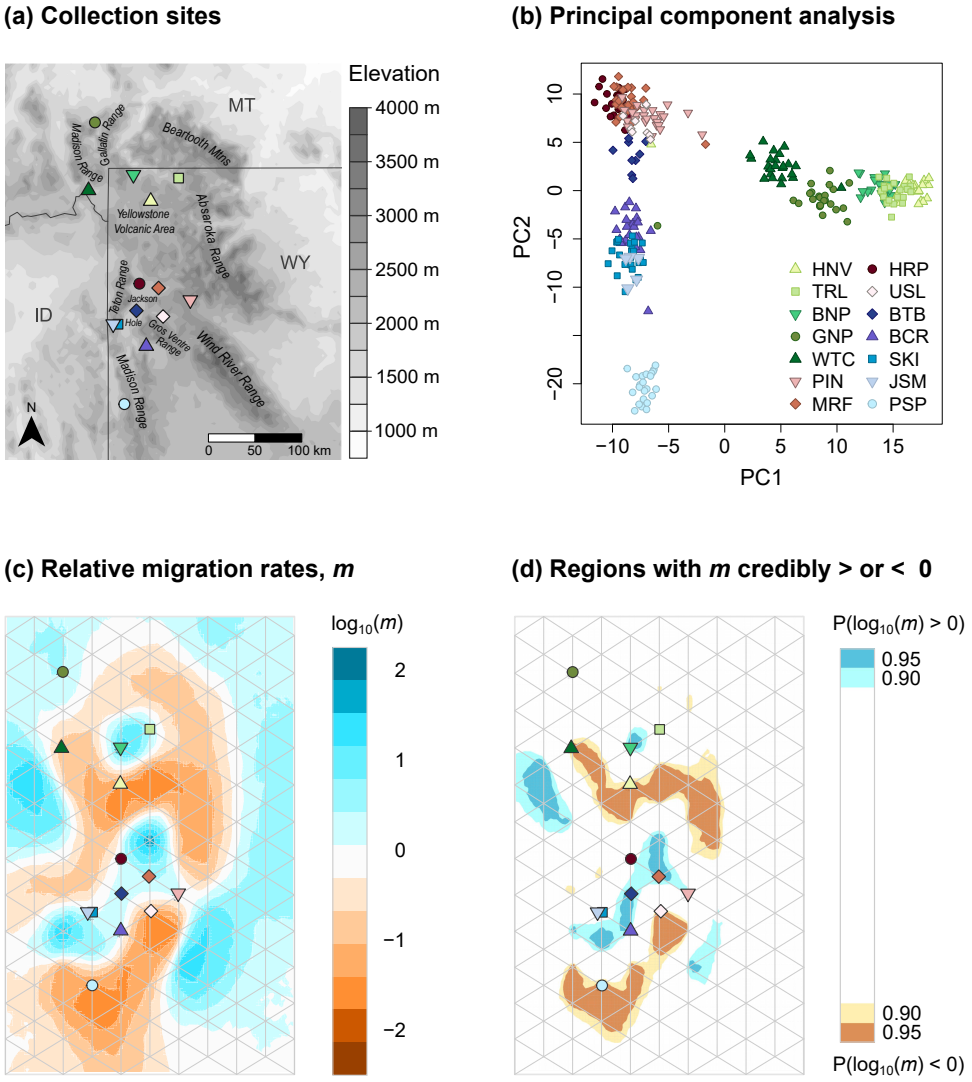

Figure 1: (a) Map of butterfly sampling locations. Each sampling site is depicted as a colored point, the corresponding key for which is shown in panel (b). Elevation contours (in meters) are shown in gray, and major mountain ranges and valley regions within *C. haydenii's* range are labeled where they occur. (b) Principal component analysis of genotype estimates from ENTROPY for the 9313 SNPs. (c) Map of relative migration rates across *C. haydenii's* range as estimated by EEMS from SNP data. Areas with estimated migration rates lower than expected under isolation by distance (IBD) alone are shown in orange, and those with migration rates higher than expected under IBD are shown in blue. Because EEMS assigns individuals to the nearest vertex on a triangular grid, the locations of populations in the EEMS model do not correspond perfectly to the sampling locations on the geographic map shown in panel (a). (d) Geographic regions with relative migration rates credibly greater or less than zero.

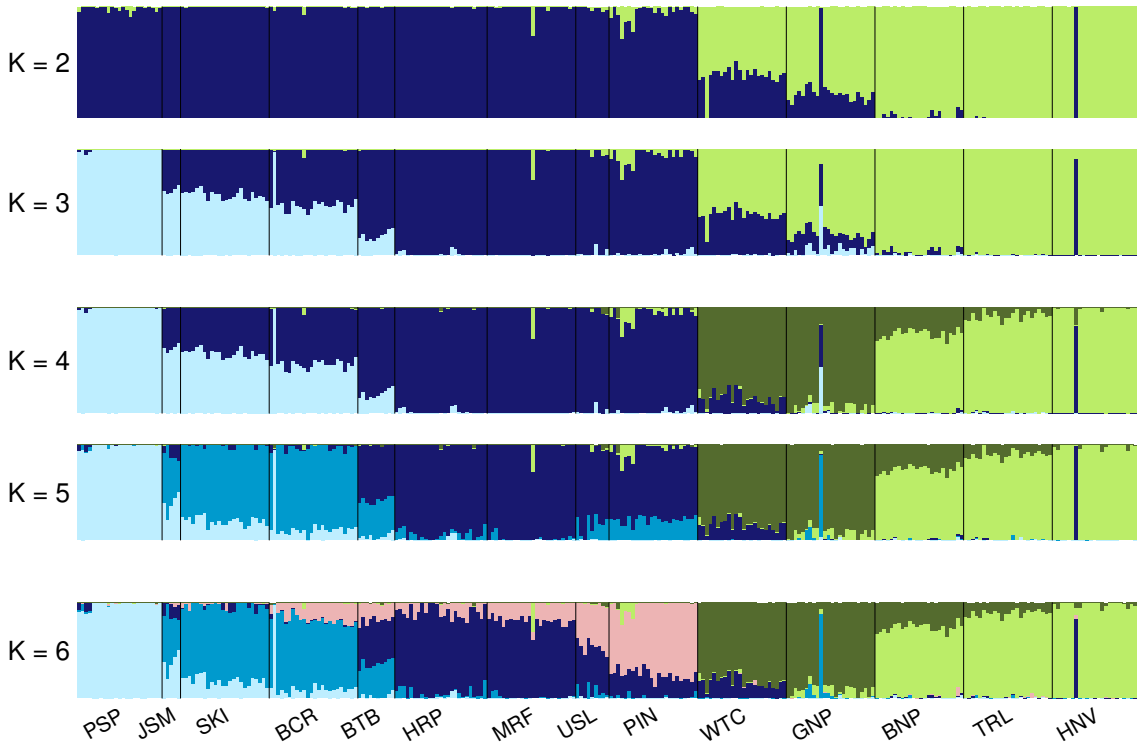

Figure 2: Estimated admixture proportions assuming individuals were sampled from $K = 2$ through $K = 6$ hypothetical source populations. Each vertical segment on the barplot represents the estimated ancestry of an individual butterfly, with the proportion of each color in the segment representing the proportion of that butterfly's genome estimated to have been inherited from each of the K putative source populations. Individuals are grouped along the x-axis by population, with populations demarcated by vertical black bars.

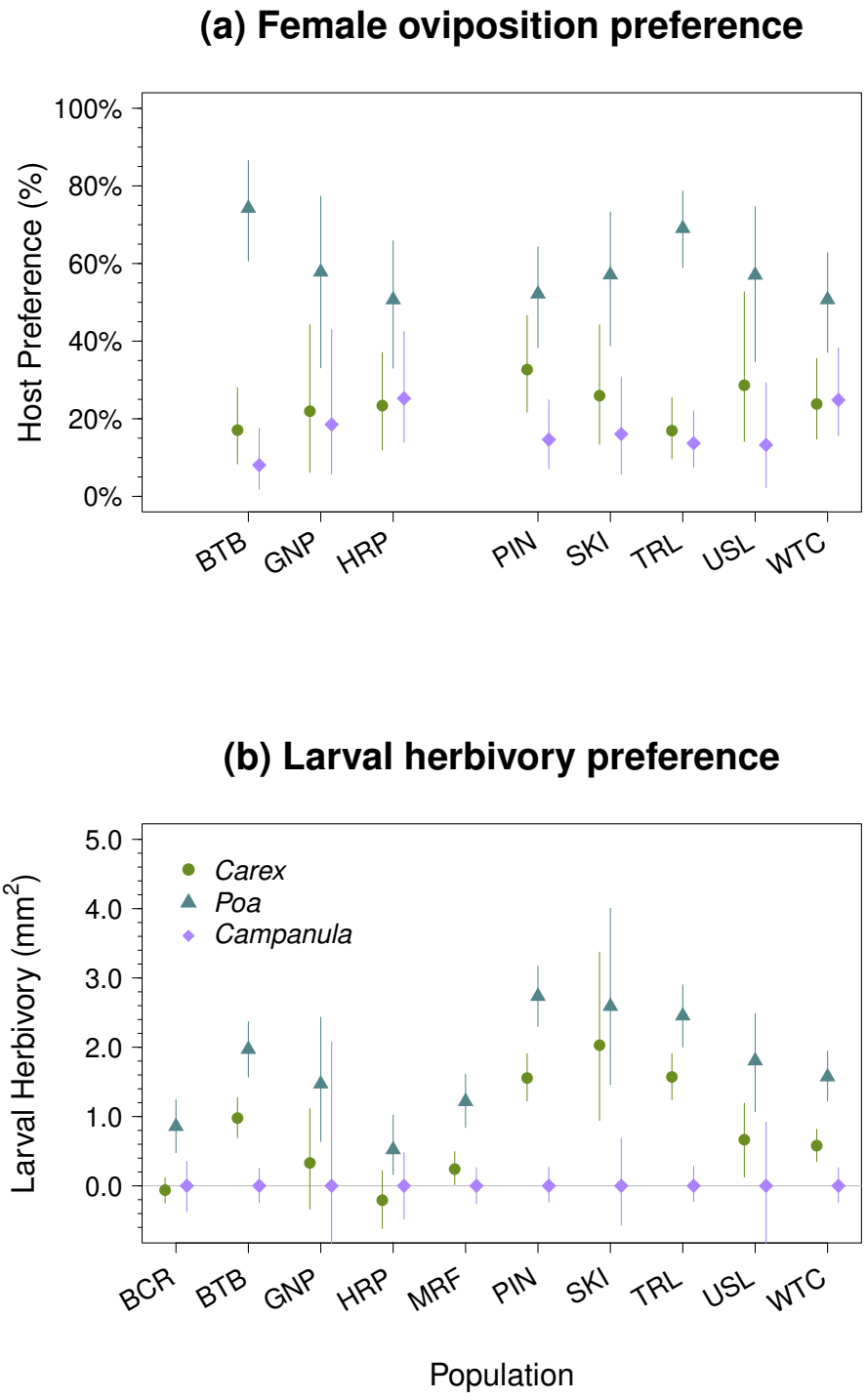

Figure 3: (a) Oviposition preference for female *C. haydenii* from 8 of our sampling sites. (b) Differences in larval herbivory across hosts for each population assayed. The expected total leaf tissue consumption for a caterpillar from a given population is shown on the y-axis. Leaves offered to larvae during the hervivory assays had a mean surface area of 15.7 mm^2.

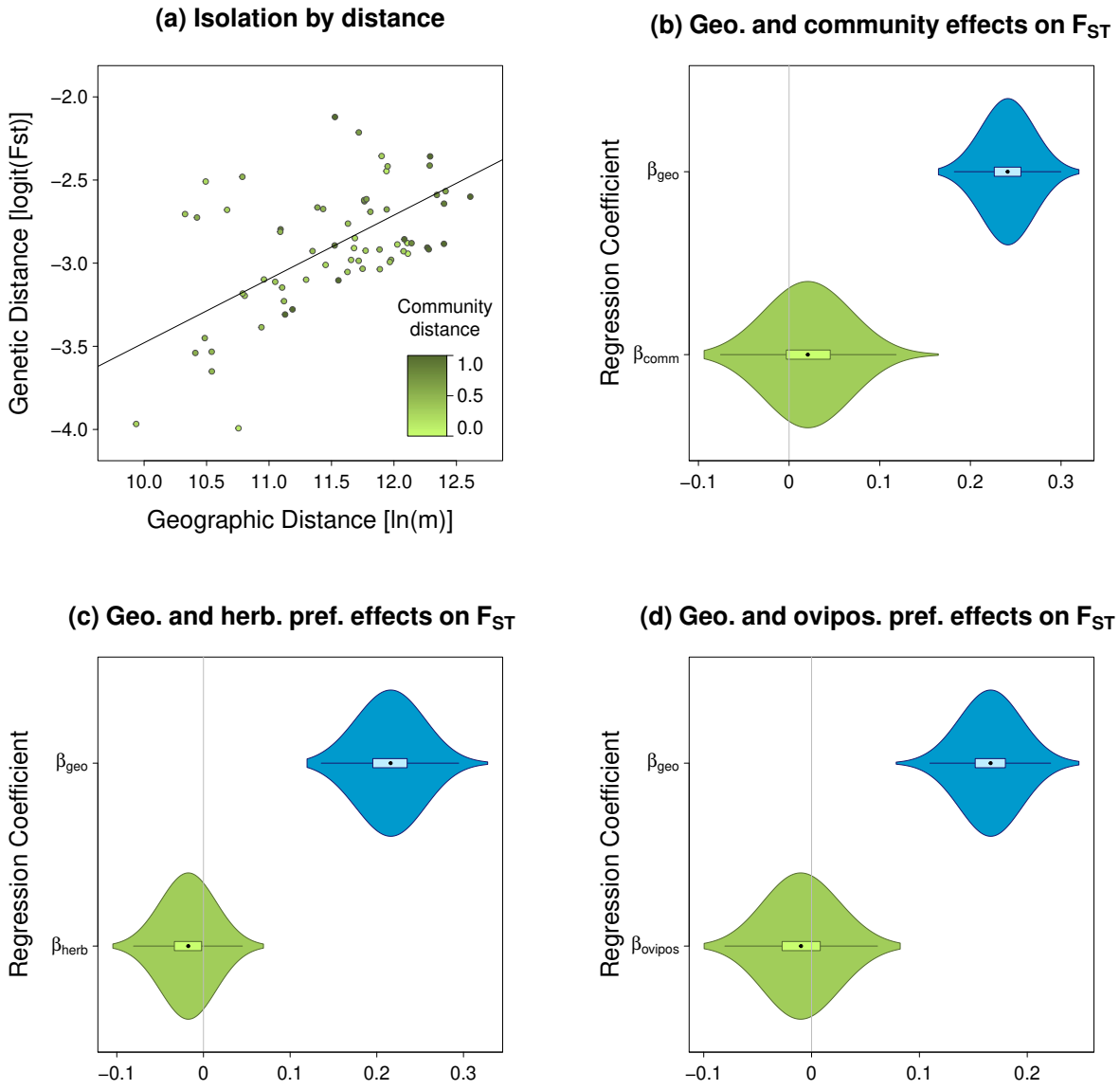

Figure 4: (a) shows the linear relationship between genetic distance (as logit F_{ST}) vs. geographic distance (ln[meters]) modeled from each pairwise combination of source populations except BTB and JSM. The color of each point on the scatter plot corresponds to the potential host community distance between each pair of sites, with lighter points corresponding to more similar host communities between sites, and darker points corresponding to more disparate host communities among sites. (b-d) show the posterior distributions for the regression coefficients in our Bayesian models estimating the degree to which geographic distance and either potential host community distance, larval herbivory preference, or oviposition preference for Kentucky bluegrass predict genetic distance (logit F_{ST}). Posterior distributions are presented in centered and standardized units for ease of comparison across regression coefficients.

Table 2: Watterson's θ (θ_W) and nucleotide diversity (π) with 95% bootstrap confidence intervals.

Population	θ_w	π
BCR	0.00348 (0.00347, 0.00349)	0.00333 (0.00331, 0.00334)
BNP	0.00360 (0.00359, 0.00362)	0.00332 (0.00329, 0.00334)
BTB	0.00311 (0.00310, 0.00313)	0.00312 (0.00309, 0.00314)
GNP	0.00347 (0.00345, 0.00348)	0.00327 (0.00325, 0.00329)
HNV	0.00301 (0.00299, 0.00302)	0.00298 (0.00296, 0.00300)
HRP	0.00335 (0.00334, 0.00337)	0.00323 (0.00321, 0.00325)
JSM	0.00280 (0.00277, 0.00283)	0.00284 (0.00281, 0.00288)
MRF	0.00353 (0.00351, 0.00354)	0.00338 (0.00336, 0.00340)
PIN	0.00330 (0.00329, 0.00331)	0.00327 (0.00325, 0.00329)
PSP	0.00327 (0.00326, 0.00329)	0.00319 (0.00317, 0.00321)
SKI	0.00339 (0.00338, 0.00341)	0.00313 (0.00311, 0.00315)
TRL	0.00349 (0.00348, 0.00350)	0.00321 (0.00319, 0.00322)
USL	0.00324 (0.00322, 0.00326)	0.00344 (0.00342, 0.00347)
WTC	0.00330 (0.00329, 0.00331)	0.00308 (0.00306, 0.00310)

Table 3: Pairwise F_{ST} values calculated from EEMS genotype estimates and geographic distances between sampling locations. Pairwise F_{ST} values are shown in the lower triangle, while geographic distances between sampling locations are shown in the upper triangle in units of km.

Population	BCR	BNP	BTB	GNP	HNV	HRP	JSM
BCR		181.9	38.9	242.8	153.6	66.5	42.1
BNP	0.0501		144.0	68.1	33.2	115.5	159.5
BTB	0.0453	0.0556		204.0	117.0	28.8	28.4
GNP	0.053	0.0353	0.0605		101.3	176.9	214.4
HNV	0.0644	0.0282	0.072	0.0525		88.2	136.0
HRP	0.0412	0.0483	0.0384	0.0544	0.0651		51.0
JSM	0.0705	0.0869	0.0849	0.0889	0.1047	0.0787	
MRF	0.0426	0.0481	0.0385	0.0532	0.0645	0.0186	0.0815
PIN	0.0381	0.0458	0.0433	0.0518	0.0594	0.0327	0.0829
PSP	0.0575	0.0665	0.0734	0.0691	0.0822	0.0674	0.0795
SKI	0.0253	0.0483	0.0421	0.0513	0.0635	0.0398	0.0609
TRL	0.0532	0.0181	0.0629	0.043	0.0284	0.0547	0.0926
USL	0.0752	0.0797	0.075	0.0864	0.0985	0.0642	0.1191
WTC	0.0508	0.0393	0.0565	0.0363	0.0567	0.0451	0.0865

Table 3: *(continued)*

Population	MRF	PIN	PSP	SKI	TRL	USL	WTC
BCR	62.9	67.5	65.7	37.9	181.1	36.1	175.7
BNP	122.7	145.4	243.0	159.1	46.9	153.3	49.3
BTB	33.5	58.0	99.7	24.1	147.2	28.9	136.9
GNP	187.3	212.6	299.8	215.0	104.4	217.4	72.3
HNV	92.3	112.6	216.7	134.8	37.8	122.8	65.5
HRP	20.7	56.3	128.5	48.5	119.1	42.8	112.2
JSM	61.3	85.4	85.5	5.1	169.1	53.4	143.9
MRF		35.8	128.2	57.5	118.2	30.6	126.6
PIN	0.0308		130.6	80.6	129.7	33.6	157.5
PSP	0.0678	0.0682		84.8	246.0	101.5	229.4
SKI	0.0432	0.0431	0.0508		167.2	48.4	145.1
TRL	0.0517	0.0509	0.0713	0.0528		147.5	94.2
USL	0.0627	0.0615	0.1071	0.0772	0.0866		154.7
WTC	0.0459	0.0477	0.0698	0.0513	0.047	0.0819	