Annual Report Vol. 40 (2017)

Spatio-temporal ecological and evolutionary dynamics in natural butterfly populations

Zachariah Gompert* and Lauren Lucas

Department of Biology, Utah State University, Logan, UT
***Author for correspondence: zach.gompert@usu.edu**

Abstract Long term studies of wild populations indicate that natural selection can cause rapid and dramatic changes in traits, with spatial and temporal variation in the strength of selection a critical driver of genetic variation in natural populations. In 2012, we began a long term study of genome-wide molecular evolution in populations of the butterfly *Lycaeides idas* in the Greater Yellowstone Area (GYA). We aimed to quantify the role of environment-dependent selection on evolution in these populations. Building on previous work, in 2017 we collected new samples, incorporated distance sampling, and surveyed the insect community at each site. We also defined the habitat boundary at a new, eleventh site. Our preliminary analyses suggest that both genetic drift and selection are important drivers in this system.

Introduction

The study of evolution in natural populations has advanced our understanding of the origin and maintenance of biological diversity. For example, long term studies of wild populations indicate that natural selection can cause rapid and dramatic changes in traits, but that in some cases these evolutionary changes are quickly reversed when periodic variation in weather patterns or the biotic environment cause the optimal trait value to change (e.g. Reznick et al., 1997; Grant and Grant, 2002). In fact, spatial and temporal variation in the strength and nature of natural selection could explain the high levels of genetic variation found in many natural populations (Gillespie, 1994; Siepielski et al., 2009). Long term studies of evolution in the wild could also be informative for biodiversity conservation and resource management, because, for example, data on short term evolutionary responses to annual fluctuations in temperature or rainfall could be used to predict longer term evolution in response to directional climate change. Most previous research on evolution in the wild has considered one or a few observable traits or genes (e.g. Kapan, 2001; Grant and Grant, 2002; Barrett et al., 2008). We believe that more general conclusions regarding the rate and causes of evolutionary change in the wild and selection's contribution to the maintenance of genetic variation could be obtained by studying genome-wide molecular evolution in a suite of natural populations. Thus, in 2012 we began a long term study of genome-wide molecular evolution in a series of natural butterfly populations in the Greater Yellowstone Area (GYA). This study will allow us to quantify the contribution of environment-dependent natural selection to evolution in these butterfly populations and determine whether selection consistently favors the same alleles across space and through time.

The focal species, *Lycaeides idas*, is one of five nominal species of *Lycaeides* butterflies that occur in North America (Figure 1; Nabokov, 1949; Guppy and Shepard, 2001; Gompert et al., 2006). These species

are descended from one or more Eurasian ancestors that colonized North America about 2.4 million year ago (Vila et al., 2011). *Lycaeides idas* hybridizes with a second species, *L. melissa*, in the GYA (Gompert et al., 2010, 2012). *Lycaeides idas* is a holarctic species that is found in Alaska, Canada, and the central and northern Rocky Mountains of the contiguous USA (Scott, 1986). *Lycaeides idas* is univoltine and adults generally fly from mid-July to early August. In the GYA *L. idas* populations often occupy mesic forest and montane habitat at elevations ranging from 2000-3500 m above sea level. Most populations of *L. idas* in the GYA feed on *Astragalus miser* as larvae, but some populations feed on other native legumes (most notably, other species of *Astragalus*, *Lupinus* and *Hedysarum*; Gompert et al., 2010). We selected *L. idas* as the focal species for this study because of our experience with this species, extensive data on the location and natural history of *L. idas* populations, the availability of genomic resources for this species, and several key aspects of this species's natural history (e.g., *L. idas* have non-overlapping generations with one generation per year, well-defined populations, and modest genome sizes, and *L. idas* are found in various habitats that might experience different environment-dependent selection pressures).

The specific goals of this study are to: (i) quantify genetic variation and molecular evolution in *L. idas* and their relationship with population size, insect community, and environmental variation across space (i.e., different populations) and through time (i.e., from generation to generation) and (ii) test the hypothesis that the nature and strength of environment-dependent selection varies among populations and over generations and that this variation is sufficiently large to contribute to the maintenance of genetic variation in *L. idas*. This report documents the results from the sixth year of this long term study. The first year (2012) was a pilot study in which we collected *L. idas* for DNA sequencing and tested the distance sampling technique to estimate population sizes. In our second year (2013) we collected *L. idas* and started distance sampling at four populations. In 2014 we collected *L. idas*, and used distance sampling at ten populations. In 2015 we collected *L. idas* from our ten focal populations. In 2016, we collected *L. idas*, attempted distance sampling at all ten populations, and tested our insect community survey method. In 2017, we collected samples, used distance sampling and surveyed the insect community (with a refined method, see below) at each site. We also defined the habitat boundary at a new, eleventh site, BLD (see below for methods).

Figure 1. *L. idas* mating pair at site PSP (see Figure 2) in the Salt River Range, WY in 2017.

Methods

We collected 484 specimens from the eleven populations involved in this study, between July 7-August 11, 2017 (Figure 2, Table 1). Specifically, we collected 39 males and 11 females from Bull Creek (BCR), 9 males and 2 females from Bald Mt. (BLD), 37 males and 13 females from Bunsen Peak (BNP), 29 males and 22 females from Blacktail Butte (BTB), 37 males and 19 females from Garnet Peak (GNP), 40 males and 10 females from Hayden Valley (HNV), 16 males and 7 females from Mt. Randolf (MRF), 36 males and 14 females from Periodic Springs (PSP), 16 males and 13 females from Rendezvous Mountain (RNV), 35 males and 18 females from Ski Lake (SKI), and 15 males and 45 females from Upper Slide Lake (USL). BNP and HNV are within Yellowstone National Park, and BTB and RNV are in Grand Teton National Park. We were unable to collect our target 50 individuals from a few of our populations: BLD, MRF, and RNV. We visited them either too late (BLD, MRF) or too early (RNV) in the season. We store these whole

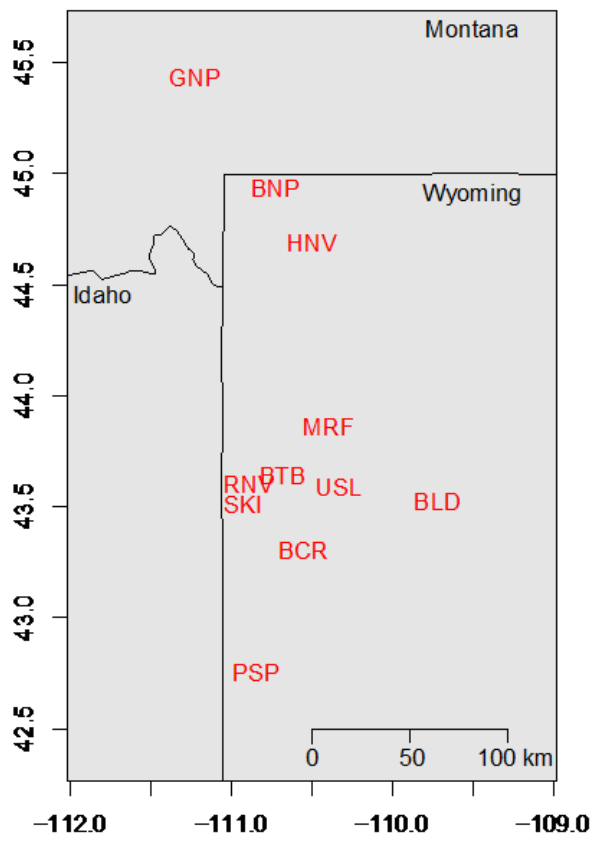

Figure 2. Map of the eleven *L. idas* populations in the GYA involved in this long term study.

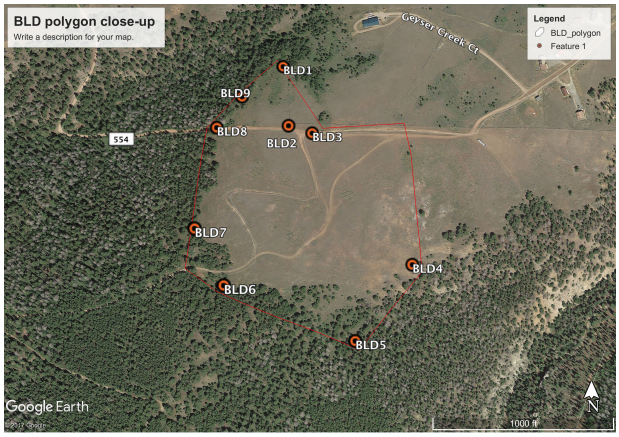

Figure 3. Habitat boundary for a new population, BLD.

adult butterflies at -80° C until DNA extraction and sequencing.

We identified a population to add to this long-term project, BLD (Figure 2 & 3). We have visited this population since 2013 for a different project. In 2017 we defined the habitat boundaries by noting GPS coordinates along the perimeter of the highest *L. idas* density in the area. We used Google Earth and the coordinates to draw a polygon around this habitat for future use in distance sampling analysis (Figure 3).

We used a distance sampling protocol to estimate *L. idas* adult population sizes at nine populations (all but RNV and BLD). Distance sampling involves counting individuals and recording their distance from a transect line or point (Buckland et al., 2001). This distance information is used to estimate a detection function that accounts for imperfect detection away from the transect line. We performed the distance sampling method one or two times per *L. idas* population over the course of the four weeks. For each population we randomly chose ten or fewer random points within a defined area of suitable habitat (we identified suitable habitat from ground surveys and satellite images). At each of these points, four trained observers (ZG, LKL, one USU Biology graduate student, Amy Springer, and one USU undergraduate student, Megan Jamison) walked an approximately 100-meter transect and: 1) counted the *L. idas* we saw along the way, recorded the sex and measured their distance on and from the transect line, and 2) quantified the abundance of butterfly host plant. We recorded a 0, 1 or 2 to denote whether there were no butterfly host-plants, less than 50% of the ground cover was host-plants, or more than 50% of the ground cover was host-plants within a meter of each transect line, respectively (Table 1). The host-plant species recorded depended on the population: *Astragalus miser* (BCR, BTB, MRF, HNV, BNP, GNP, SKI, USL), *Astragalus bisulcatus* (USL), *Lupinus* sp. (PSP) or *Hedysarum* sp. (RNV, SKI). We only performed distance sampling between 10:00 am and 3:00 pm under sunny or partly sunny skies.

We estimated population densities (adult butterflies per square kilometer) using the *distsamp* function in the *unmarked R* package. We binned the detection distances of butterflies into 1 meter bins prior to analysis (e.g., 0 to 1 m, 1 to 2 m, etc.). We used a half-normal detection function and estimated the detection function and density model parameters using

maximum likelihood (Royle et al., 2004). This model assumes the latent transect-level abundance distribution is Poisson and that the detection process is multinomial with a different detection probability for each distance class or bin. We then estimated population size by first multiplying density by the area of habitat (km^2) and then by three because adult *L. idas* live for about a week but the population flies for about three weeks.

To preliminarily explore whether differences in population size across space (populations) can be explained by climate, we used 19 weather variables averaged over 1950-2000 (source: http://www.worldclim.org/bioclim), summarized as one variable via a Principal Component Analysis (PCA) using the *prcomp* function in *R*.

We conducted an insect community survey at several locations within the habitat boundary of each of nine populations (all but RNV and BLD), typically at the start or end of the census population transects (10 at BTB, 12 at PSP, 13 at BCR, 11 at USL, 7 at GNP, 8 at BNP, 4 at HNV, 15 at MRF, 10 at SKI). The survey consisted of approximately 3 sweeps of *L. idas* host plant with a sweep net. Insects in the sweep net were identified to family when possible (otherwise order or suborder), and the number of individuals of each family were recorded. The main difference between our 2016 and 2017 methods was we emphasize the presence of insects that are known to positively or negatively interact with *Lycaenid* butterflies (however, we recorded the presence of others for our records): *Opiliones, Araneae, Membacidae, Cicadellidae, Ceropidae, Aphididae, Miridae, Reduviidae, Nabidae, Penta/Scutelleridae, Formicidae, Ichneumonidae, Braconidae, Chalcidoidea, Acrididae/Gryllidae,* and *Hemer/Chrysopidae*. These data were used in a principal component analysis (PCA) to characterize differences in insect community across sites. We used the *prcomp* function in *R* to perform this PCA.

We completed a genotyping-by-sequencing (GBS) survey of genetic variation in *Lycaeides* (as in, Gompert et al., 2014) from samples collected from eight populations sampled in 2013 and 2015. Genetic change in these sites across years was used to estimate contemporary variance effective populations sizes (as in, Gompert and Messina, 2016). During 2017-2018, we have been preparing GBS libraries for the 2017 collections made at 10 of our sites (all but BLD), and for our 2013-2014 HNV and SKI samples that were excluded in the previous round of sequencing. These sequence results will be analyzed and reported in our 2018 report.

Results

In Table 1 we include our population size estimates from distance sampling in summers 2013-2014 and 2016-2017, as well as average host-plant abundance collected during 2017. For data collected in 2017, *L. idas* abundances were high enough at only four sites to analyze distance sampling data. When comparing the new 2017 estimates to previous years' data, we observed that all estimates were at least slightly lower than they have been for previous years. The range of host-plant abundance across sites was 0.2 to 0.9, with the highest abundance at SKI and BNP and the lowest at HNV, which is similar to 2016 abundance estimates. The climate variable (PC1 from the PCA, which explained 41% of the variance in climate among sites) ranged from -3.4 to 5.8 across sites. Negative numbers represent hotter and drier climates, whereas positive values represent colder and wetter climates. We found that PSP (-3.4), BCR (-2.9) and BLD (-2.9) were the hottest/driest. The coldest and wettest were RNV at 5.8 and GNP at 2.5 (Table 1).

The PCA performed on the insect community data revealed differences among sites (Figure 4). Principal component 1 (PC1) explained 27% of the variance in the dataset, and PC2 explained 16% of the variance. PC1 was a bit different for PSP and BNP, which are on opposite sides of the latitudinal gradient, and BTB was different than the other sites for PC2.

Contemporary estimates of effective population sizes (Figure 5), based on genetic data, varied by an order of magnitude more than and were uncorrelated with genetic diversity levels (coefficient of variation = 0.196 vs 0.014 for expected heterozygosity). These

Pop	2013 size	2014 size	2016 size	2017 size	2016 host plant	Climate PC1 score
BTB	1838.7	1978.5	2763.3	1155.3	0.7	-0.7
BCR	2382	1241.7	NA	NA	0.5	-2.9
BNP	633.9	1273.2	NA	NA	0.9	1.0
GNP	1119.9	1024.5	NA	343.0	0.4	2.5
HNV	NA	5291.4	NA	NA	0.2	1.4
MRF	NA	977.7	NA	NA	0.4	-0.9
PSP	NA	366.6	NA	354.2	0.6	-3.4
RNV	NA	NA	NA	NA	NA	5.8
SKI	NA	1348.8	1242.2	694.4	0.8	1.9
USL	NA	1708.2	2927.1	NA	0.5	-1.9
BLD	NA	NA	NA	NA	NA	-2.9

Table 1. Population abbreviations ("pop"), population size estimates via distance sampling in 2013-14 and 2016-17, average host-plant abundance (estimated in 2016; "host plant"), and a representation of long term climate at each population (PC1 represented 41.9% of the variance in the original dataset).

results suggest that diversity might not reflect drift-mutation equilibrium as posited by standard neutral theory, and that alternative hypotheses, such as widespread fluctuating selection, warrant consideration.

Discussion

Based on our moderate population size estimates, we predict that both genetic drift and selection are important drivers of evolution in this system (Lynch, 2007). The comparison of population size estimates among years is potentially interesting and could reflect demographic variability across time. The difference in habitat (i.e., host-plant and insect community) and climate across populations highlights the spatial variation in this study system.

We will continue this study during the 2018 summer field season. During this and subsequent field seasons, we will collect samples and estimate population sizes at all 11 sites listed in Table 1. We plan to compare the mark-recapture method for estimating census population size to the distance sampling method we have been using at a few of our sites to validate our methods. We will also continue collecting habitat data that will be useful for fitting causal models of molecular evolution.

References

Barrett, R. D., S. M. Rogers, and D. Schluter. 2008. Natural selection on a major armor gene in threespine stickleback. Science **322**:255–257.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. Introduction to distance sampling: estimating abundance of biological populations. Oxford University Press, USA.

Gillespie, J. H. 1994. The causes of molecular evolution. Oxford University Press, USA.

Gompert, Z., J. A. Fordyce, M. L. Forister, A. M. Shapiro, and C. C. Nice. 2006. Homoploid hybrid speciation in an extreme habitat. Science **314**:1923–1925.

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

Gompert, Z., L. K. Lucas, J. A. Fordyce, M. L. Forister, and C. C.

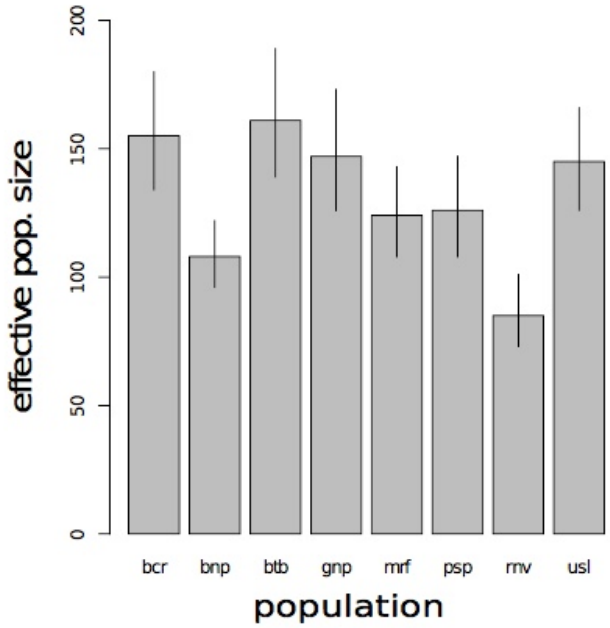

Figure 5. Variance in effective population size estimates from 8 focal populations based on genome-wide measures of allele frequency changes.

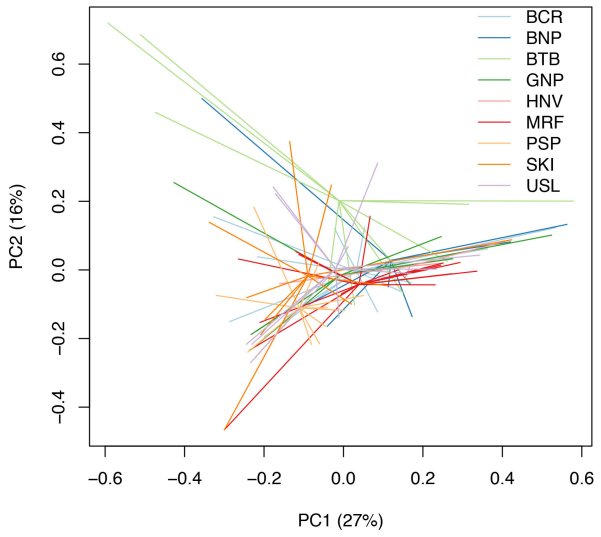

Figure 4. Differences in the abundances of major insect groups across nine *L. idas* sites, as summarized by a PCA.

Nice. 2010. Secondary contact between *Lycaeides idas* and *L. melissa* in the Rocky Mountains: extensive admixture and a patchy hybrid zone. Molecular Ecology **19**:3171–3192.

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

Gompert, Z., and F. J. Messina. 2016. Genomic evidence that resource-based trade-offs limit host-range expansion in a seed beetle. Evolution **70**:1249–1264.

Grant, P. R., and B. R. Grant. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. Science **296**:707–711.

Guppy, C., and J. Shepard. 2001. Butterflies of British Columbia. UBC Press.

Kapan, D. D. 2001. Three-butterfly system provides a field test of Müllerian mimicry. Nature **409**:338–340.

Lynch, M. 2007. The origins of genome architecture. Sinauer Associates Sunderland, MA.

Nabokov, V. 1949. The nearctic members of *Lycaeides* Hübner (Lycaenidae, Lepidoptera). Bulletin of the Museum of Comparative Zoology **101**:479–541.

Reznick, D. N., F. H. Shaw, F. H. Rodd, and R. G. Shaw. 1997. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). Science **275**:1934–1937.

Royle, J. A., D. K. Dawson, and S. Bates. 2004. Modeling abundance effects in distance sampling. Ecology **85**:1591–1597.

Scott, J. A. 1986. The butterflies of North America: a natural history and field guide. Stanford University Press.

Siepielski, A. M., J. D. DiBattista, and S. M. Carlson. 2009. It's about time: the temporal dynamics of phenotypic selection in the wild. Ecology Letters **12**:1261–1276.

Vila, R., C. D. Bell, R. Macniven, B. Goldman-Huertas, R. H. Ree, C. R. Marshall, Z. Balint, K. Johnson, D. Benyamini, and N. E. Pierce. 2011. Phylogeny and palaeoecology of *Polyommatus* blue butterflies show Beringia was a climate-regulated gateway to the New World. Proceedings of the Royal Society B: Biological Sciences **278**:2737–2744.