

PHYLOGEOGRAPHY AND DIVERSIFICATION OF THE GREENISH BLUE
BUTTERFLY (*PLEBEJUS SAEPIOLUS*) IN WESTERN NORTH AMERICA

THESIS

Presented to the Graduate Council of
Texas State University-San Marcos
in Partial Fulfillment
of the Requirements

for the Degree

Master of SCIENCE

by

Desirae M. Weyland, B.S., B.A.

San Marcos, Texas
December 2013

PHYLOGEOGRAPHY AND DIVERSIFICATION OF THE GREENISH BLUE
BUTTERFLY (*PLEBEJUS SAEPIOLUS*) IN WESTERN NORTH AMERICA

Committee Members Approved:

Chris Nice, Chair

James Ott

Noland Martin

Approved:

Andrea Golato
Dean of the Graduate College

COPYRIGHT

by

Desirae M. Weyland

2013

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

This work is protected by the Copyright Laws of the United States (Public Law 94-553, section 107). Consistent with fair use as defined in the Copyright Laws, brief quotations from this material are allowed with proper acknowledgement. Use of this material for financial gain without the author's express written permission is not allowed.

Duplication Permission

As the copyright holder of this work I, Desirae M. Weyland, authorize duplication of this work, in whole or in part, for educational or scholarly purposes only.

ACKNOWLEDGEMENTS

I would like to thank Chris Nice, my advisor, for accepting me as a student, and guiding me and supporting me in this and other research. Additionally, I would like to thank my thesis committee, Jim Ott and Noland Martin, for their support, input and revisions of the manuscript. Also, thank you to the Nice lab group and several graduate students of my cohort, especially Kate Bell, for her indispensable support, guidance and friendship.

This manuscript was submitted on August 5, 2013.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF TABLES	vii
LIST OF FIGURES	viii
ABSTRACT	ix
CHAPTER	
I. THESIS	1
Introduction	1
Methods and Materials	6
Results	11
Discussion.....	15
APPENDIX	35
LITERATURE CITED.....	40

LIST OF TABLES

Table	Page
1. Sampling Locations and Designations for <i>Plebejus saepiolus</i>	20
2. Pairwise Φ_{ST} Values for All Sampling Location Comparisons	21
3. Tajima's D and Fu's F Values for All Sampling Locations	22
4. SAMOVA Results	23
5. Results of all AMOVA Analyses	24
6. Results of One-Way ANOVAs for Principle Component 1	25
7. Tukey's HSD Comparisons by Sampling Locations and Groups	26

LIST OF FIGURES

Figure	Page
1. Map of Sampling Locations and Range Map for <i>Plebejus saepiolus</i>	27
2. Wing Spot Measurements and Examples from Differentiated Populations	28
3. Haplotype Networks by Population and Groups	29
4. Graph Showing Optimized Φ_{ST} by Number of Groups (k)	30
5. Map of Population Groupings Based on SAMOVA Results	31
6. Non-metric Multi Dimensional Scaling (NMDS) plot of Pairwise F_{ST} 's for All Populations	32
7. Graphs of Genetic Distance by Geographic (Orthodromic) Distance	33
8. Principle Components Analysis (PCA) Biplots by Populations and Groups	34

ABSTRACT

PHYLOGEOGRAPHY AND DIVERSIFICATION OF THE GREENISH BLUE BUTTERFLY (*PLEBEJUS SAEPIOLUS*) IN WESTERN NORTH AMERICA

By

Desirae M. Weyland, B.S., B.A.

Texas State University-San Marcos

December 2013

SUPERVISING PROFESSOR: CHRIS NICE

Investigation of an organism's biogeography is an important first step in understanding evolutionary processes. Often large-scale climatic or geographic events, like climate cycling of the Pleistocene, can influence dispersal patterns and create a disjunct and complex distribution in a species range. The Greenish Blue Butterfly, *Plebejus saepiolus* (Lycaenidae, Polyomattini), is a species which is widespread in North America, but populations are restricted to montane habitat. Here we use molecular genetic data and morphometric analyses of wing pattern variation to address the following questions (i) How has fragmentation and divergence occurred across populations based on current genetic variation in mitochondrial DNA based on phylogeographical hypotheses? (ii) How and where did colonization of populations in *P. saepiolus* occur across its geographic range? (iii) Does wing pattern variation parallel molecular differentiation, and agree with current subspecific designations within *P.*

saepiolus? Overall, 31 unique haplotypes were identified in *P. saepiolus* with moderate isolation by distance, but no evidence of rapid expansion. Multiple phylogeographic models can characterize the complex genetic pattern in *P. saepiolus*. Genetic variation partitioned by mountain range with a few exceptions, however wing pattern variation does not parallel the genetic variation, and taxonomic designations should be reevaluated.

CHAPTER I

THESIS

I. Introduction

Investigation of an organism's biogeography is an important first step in understanding evolutionary processes (Avice 1994, Hewitt 2000, Avice 2000, Hewitt 2004, Nice et al. 2005, Lomolino et al. 2010). Often large-scale climatic or geographic events, like climate cycling of the Pleistocene, can influence dispersal patterns and create a disjunct and complex distribution in a species range (Avice et al. 1979, Avice 1994, Hartl and Clark 1997, Avice 2000). Organisms with large geographic ranges and polymorphic populations make good study species for testing phylogeographic hypotheses (Avice et al. 1979, Nice et al. 2005). By examining current population distributions and patterns of genetic, morphological and ecological variation, we can discover what temporal and spatial processes contribute to diversification and influence species histories.

The Greenish Blue Butterfly, *Plebejus saepiolus* (Lycaenidae, Polyommastini), is a species which is widespread in North America, but populations are restricted to montane habitat (Brock and Kaufman 2003, Glassberg 2001). As with most lycaenid butterflies (Austin and Murphy 1987, Nice et al. 2005), this species cannot survive far from host and nectar sources and is not a strong flier. Populations of *P. saepiolus* occur in mountain ranges on either side of vast areas of unsuitable habitat (Brock and Kaufman 2003) that

should prevent dispersal between ranges in a short lived, non-vagile species. At least 12 subspecies (4 proposed in 1998 by Emmel & Brown) have been identified based on either differences in wing pattern morphology, regional isolation, or both. Three of these subspecies, the Insular or Coastal Blue (*P. saepiolus littoralis*), the Island Blue (*P. s. insularis*) and the San Gabriel Blue (*P. s. aureolus*) are considered imperiled or extinct within their range (Pyle 2002, Heron 2007). *Plebejus s. littoralis* occurs in coastal northern California (Lake Earl, Del Norte County) and southern coastal Oregon (Curry and Coos Counties). As it's name suggests, this subspecies is found at low elevations (roughly 50ft above sea level) and is often found near dune habitats, making it an exception to the mostly montane distribution of the species. This subspecies is identified by enlarged white halos surrounding wings spots, as well as less black spotting overall (Pyle 2002). *Plebejus s. insularis* is protected in British Columbia, and is only documented to occur on Vancouver Island in Canada (Heron 2007). Individuals of this subspecies have not been observed since 1979, however there is debate about whether the subspecies is distinct from other populations through the western United States (Heron 2007). Last, *P. s. aureolus* occurred in the San Gabriel Mountains of the Angeles National Forest in southern California, but is now thought to be extinct due to draining of Big Pine recreational area (Murphy 1990, Stephenson and Calcarone 1999). The subspecies has not been observed for over a decade and, like the Island Blue, this subspecies was not described as having distinct morphological traits (Murphy 1990, Stephenson and Calcarone 1999).

Plebejus s. albomontanus, the proposed subspecies in the White Mountains, may also be distinct from other subspecies, or represent a point where Sierran and Rocky

Mountain lineages meet. Several distinct wing morphologies have been observed in this location, in *P. saepiolus*, as well as in other butterfly species (CCN pers. comm.). The proposed White Mountain subspecies also occurs at an extremely high elevation compared to the majority of Rocky Mountain and Sierra Nevada populations of *P. saepiolus*. These and other populations within *P. saepiolus* may be genetically differentiated and isolated, meriting conservation consideration. Currently, the taxonomy within the species is unresolved and there have been no genetic or morphological analyses to support subspecies designations.

Pleistocene glaciations have had impacts on the demography and spatial distribution of several alpine and montane organisms in the United States (Pielou 1992, Schmitt 2007). Three phylogeographical models of dispersal during Pleistocene glaciation cycles may separately or concurrently explain patterns of differentiation in populations observed across mountain ranges in western North America. The expanding alpine archipelago model (Knowles 2001) predicts population expansion and large colonization events to new locations, followed by contraction during interglacial periods. This model predicts corridors developed between mountain ranges during glacial periods, allowed population expansion and large colonization events to new locations, followed by contraction during interglacial periods (Knowles 2001, DeChaine and Martin 2004, Schoville and Roderick 2009, Schoville et al. 2011). This model may be consistent with what has occurred in the expansion of *P. saepiolus* across the Great Basin ranges during glacial periods. The alpine archipelago refuge model (Schmidt 2007), on the other hand, proposes small dispersal events occurred from source populations during glacial periods, and expansion occurred during interglacial periods (Schmidt 2007, Schoville and

Roderick 2010, Schoville et al. 2011). A pattern of isolation by distance would support this scenario (Schoville et al. 2011). Last, the ancestral radiation and fragmentation model (Knowles 2001) proposes a rapid radiation occurred from a large ancestral population that was subsequently fragmented and isolated (Knowles 2001, Schoville et al. 2011). Patterns of genetic variation, with little geographic structuring, and evidence of past bottlenecks would be consistent with this model. Each of these models is characterized by specific genetic patterning across the landscape; however, these models are not mutually exclusive and events predicted by these models may have occurred concurrently or sequentially on populations leaving signatures of multiple models.

In other phylogeographic studies of montane organisms of North America, dispersal has originated in the Rocky Mountains with newer populations established in the western mountain ranges (Siskyou, Sierra Nevada, Coast Range)(Nice and Shapiro 2005, Schoville and Roderick 2009). If colonization occurred by means of several founder events, increased haplotype diversity should be observed in older or source populations, and should be lower in populations founded more recently or farther along the dispersal route (Cann et al. 1987, Templeton 1998, Nice and Shapiro 2001, Petit 2011). Often dispersal of populations occurs leptokurtically, with a few early dispersers followed by a large colonization event and few late dispersers, creating a wave like pattern of colonization across the landscape (Hewitt 1996, Ibrahim et al. 1996). In the case of leptokurtic dispersal, earlier and later colonized populations should have decreased genetic diversity, because they have fewer founders (Ibrahim et al. 1996, Hewitt 1996, Petit 2011). Additionally, it is more likely that range expansion occurred rapidly, rather than having constant migrations over an extended period of time (Petit

2011). When dispersal is leptokurtic, it results in a patchy distribution with clustered genotypes, which persist for hundreds of generations (Ibrahim et al. 1996) For example, if a single rapid migration across the Great Basin from the Rocky Mountains occurred, as has been observed in other alpine and montane organisms of western North America, we would expect that all populations west of the Rockies to be more closely related to one another than they are related to populations in the Rocky Mountain populations (Galbreath et al. 2010).

Examination of currently isolated populations may be particularly informative in inferring biogeographic histories of *P. saepiolus*. The White Mountains also represent the highest elevation site sampled, possibly further isolating it from nearby southern Sierran populations. Additionally, several other studies have found unique species and populations for the White Mountains (Billings 1973, Morefield 1992, Jennings and Elliott-Fisk 1993, Schoville et al. 2012) The White Mountains may have served as an island during colonization, and is observed to have a great amount of phenotypic variation. It is possible that this area was colonized from both the east (Rockies populations), as well as from the northwest by populations moving across the Sierra Nevada Range. Like the White Mountains, the Lake Earl site may be more isolated from neighboring populations because of its elevation. These sites may therefore exhibit more unique, and less shared haplotypes than other populations.

Although wing spot variation in *P. saepiolus* has not been previously quantified, some subspecific variation in marginal wing spot bands have been noted. It is possible that this variation in spot pattern is due to isolation and, therefore may coincide with phylogeographic patterns of variation. Additionally, evaluating both morphological data

and genetic data allows reassessment of current and proposed subspecific designations in *P. saepiolus*. This may be important since wing spot variation is currently the main character describing differences in subspecies, and may not reflect actual underlying genetic variation. For example, designation of the low elevation subspecies *P.s. littoralis* is based on diminished black spotting in marginal bands. Although wing pattern may be partially plastic, these changes may also have a genetic component. Additionally, in some Lycaenid butterflies hind wing spot pattern is shown to be important in mate discrimination (Fordyce et al. 2002), thus creating a barrier to mating between distinct populations. This may be important in redefining morphological traits that define subspecies, or potentially diverging populations.

Here we use molecular genetic data and morphometric analyses of wing pattern variation to address the following questions (i) How has fragmentation and divergence occurred across populations based on current genetic variation in mitochondrial DNA based on phylogeographical hypotheses? (ii) How and where did colonization of populations in *P. saepiolus* occur across its geographic range? (iii) Does wing pattern variation parallel molecular differentiation, and agree with current subspecific designations within *P. saepiolus*? Together these objectives allow us to assess how historical climate cycles have shaped phylogeographical structure affected population level variation in the species.

II. Methods and Materials

i. Species History and Population Sampling

Plebejus saepiolus (Lycaenidae, Polyomattini), a small (1-2 inch wingspan), blue butterfly, is one of twelve members of its genus occurring in Western North America (Scott 1992, Opler 1999, Brock and Kaufman 2003). Ancestors of North American Polyommatus blue butterflies are thought to have migrated via the Bering land bridge in five separate invasions during the Miocene-Pleistocene cooling cycles (Vila et al. 2011). *Plebejus saepiolus* diverged from individuals colonizing North America during the second invasion of Polyommatus blues, roughly 9.3 million years ago, during the late Miocene (Vila et al. 2011). The current range of *P. saepiolus* encompasses western North America from Alaska to California and has been expanding to the east in the Northern US towards New England since 1912 (Figure 1) (Scott 1992, Opler 1999). *Plebejus saepiolus* is found mainly in subalpine meadows, though a few low elevation coastal populations have been observed (Scott 1992).

Two hundred twenty-one *P. saepiolus* were collected between 1994 and 2009 from locations throughout eight regions in its range, including the Sierra Nevada, the White mountains, the Sweetwater Mountains, the Warner Mountains, the Coast Range, Siskiyou Mountains, Rocky Mountains and Alaska (Table 1). 10-16 individuals were collected from each of 16 sampling localities and were grouped according to geography into eight *a priori* regions for the analyses.

ii. Mitochondrial DNA

Total genomic DNA was extracted from a dissected portion of the thorax in all individuals using the DNAeasy Blood and Tissue Kit (Qiagen, Valencia, CA) following manufacturer protocols. The mitochondrial gene Cytochrome oxidase subunit II (COII) gene was amplified using specific primers (Pierre 5' AGA GCC TCT CCT TTA ATA

GAA CA 3' and Eva 5' GAG ACC ATT ACT TGC TTT CAG TCA TCT 3') in Polymerase Chain Reactions (Simon et al. 1994, Caterino and Sperling 1999, Runquist et al. 2012). 50ul per individual aliquots of PCR mix were prepared as follows: 32.32ul autoclaved, double-distilled water, 10.00ul of Buffer A, 1.00ul deoxyribonucleotide mix (10uM dNTPs), 2.00ul of magnesium chloride, 1.25ul of each primer (10 uM), 1.00ul of dimethyl sulfide (DMSO), 0.25ul of Taq DNA polymerase (GoTaq flexitm Promega), and 1.00ul of genomic DNA. The product was optimized for amplification under the following conditions: initially denatured for 2 minutes at 94 °C; 1 minute additional denaturation at 94 °C, 1 minute anneal at 48 °C, and 1 min extension at 72 °C repeated for 35 cycles; followed by 10 minute extension at 72 °C, and stored at 4 °C until electrophoresis. All PCR amplifications (50ul reactions) were performed on a thermocycler. PCR product was prepared for sequencing using Wizard SV gel and PCR clean-up system following standard protocols with 50ul elution. Sanger sequencing with forward amplification primer was performed at the Nevada Genomics Center, at the University of Nevada, Reno for approximately 492 base pair region of the COII mitochondrial gene on all samples. All sequencing was performed using ABI BigDye Terminator Cycle Sequencing Ready Reaction Kit v3.1 and run on the ABI3730 DNA Analyzer.

iii. Data Analysis

Sequences were checked and aligned using Geneious version 5.3.5. Indels were removed from the entire sample by deleting corresponding bases across all individuals to allow analyses of sequence data. All sequences were translated into amino acid residues to ensure accuracy of data, and to ensure no stop codons occur within the sequences.

Measures of genetic differentiation (Φ_{ST} , Φ_{CT} , and Φ_{SC}) within and among putative populations and mountain ranges (regions) were estimated with Analysis of Molecular Variance (AMOVA) in Arlequin Version 3.5 (Excoffier and Lischer 2010). Additionally, Arlequin Version 3.5 was used to identify private alleles and to calculate measures of nucleotide diversity (π) and haplotype diversity (h) at each site (Excoffier and Lischer 2010). Negative and significant Tajima's D and Fu's F values across or within regions indicate departure from a neutral drift-mutation model and test for neutrality and population expansion in *P. saepiolus*, and were calculated in Arlequin version 3.5 (Tajima 1989, Excoffier and Lischer 2010). A mismatch distribution, to identify population expansion based on Harpending's raggedness index, was performed in Arlequin Version 3.5 in addition to evaluation of Tajima's D and Fu's F values (Excoffier and Lischer 2010).

JModelTest version 2.1.3 (Darriba et al. 2012), was used to identify the HKY+I (proportion of invariant sites equal 0.8190) as the best fit model of sequence evolution based on Bayesian Information Criterion and was used to estimate sequence distances in the program PAUP* (Swofford 2002). Statistically parsimonious haplotype networks were estimated using TCS reflecting site and regional designations (Clement et al. 2000). A Spatial Analysis of Molecular Variance for $k=2-16$ was conducted using the program SAMOVA (Dupanloup et al. 2002) to find the regional grouping that maximizes genetic differentiation among groups with no *a priori* assignment of regional grouping by comparing among-region component of variance (Φ_{CT}) across values of k . Based on the observed geographic pattern, isolation by distance was tested using a Mantel's Test for correlation between pairwise genetic distances (pairwise Φ_{ST}) and pairwise orthodromic

geographic distances using R (Dray and Dufour 2007, R Core Team 2012). Additionally, construction of non-metric multidimensional scaling (NMDS) analysis of pairwise Φ_{ST} was employed to illustrate pairwise comparison of regional differentiation using R (Venables and Ripley 2002, R Core Team).

iv. Wing Morphology

Plebejus saepiolus males have iridescent blue dorsal scales, while females are brown-orange in dorsal color; both sexes possess a black forewing cell bar with a marginal black or orange band and white fringe at the distal wing margins (Scott 1992, Opler 1999, Glassberg 2001, Brock and Kaufman 2003). The ventral wing surface is grey to pale blue, characterized by several rows of black spots encompassed in white halos; the marginal and submarginal spot rows are usually as intense as the post median spot row, but submarginal spots are often elongated (Brock and Kaufman 2003, Scott 1992). Wing morphology, specifically post median and submarginal hind wing spot areas were quantified using ImageJ software. Wings of at least 6 males per collection locality were photographed under equal light conditions. Some individuals were excluded based on excessive wing wear and analyses were confined to only males. Although distinguishing characters for subspecific identification have been reported as more pronounced in females (Pyle 2002), sampled individuals are predominantly male across all collected populations. Thus, only male butterflies were used in analyses of wing pattern in order to control for any effect of sexual dimorphism. Specific measurements of wing area and area of 19 black spots were recorded for each individual (Figure 2). Spot size was corrected by dividing by wing area in order to account for allometric size differences in spots. All photos were converted to greyscale and the colors inverted in order to highlight

wing spots and wing bars. Any spots not present in an individual, but present in at least one other individual in the sample were coded as a zero. Principal Components Analysis (PCA) analyses of wing spot measurements in R (Venables and Ripley 2002, R Core Team 2012) were used to illustrate variation among individuals and populations. To compare morphological grouping to the genetic relationships between populations, principle component scores were used to compare spot variation within and among sites and regions using multiple Analyses of variance (ANOVA) in R (R Core Team 2012).

III. Results

i. mtDNA Analyses

A total of 221 individuals of *P. saepiolus* from 16 locations across the species' range were sequenced for the 492 bp region of mitochondrial gene COII, yielding 31 unique haplotypes (Figure 3). Several locations exhibited private haplotypes, with 23 of the observed haplotypes only occurring at a single site. Additionally, three sites in the Sierra Nevada (Mt. Rose, Tioga Crest and Donner Pass) were fixed for haplotype A, which is the most frequent and widespread haplotype, occurring in 10 sampled locations (haplotype A, n=83). This haplotype was present in all Sierran sites, and was not observed in any individuals from the Rocky Mountains, Alaska, Siskyou, or Lake Earl sites. Haplotype B (n=32) and haplotype H (n=27) were the next most common haplotypes (Table 1). Additionally, Haplotype B was observed in all sites in Warner Mountains, Siskyou Mountains, and Alaska. Additionally, the coastal, low elevation Lake Earl site, and the Sierra Nevada Carson Pass and Leek Springs sites were represented by this haplotype. Haplotype B was not observed in the Rocky Mountains sites, the

Sweetwater Mountains, the White Mountains, and the Plaskett Meadows site. This haplotype was extremely abundant in the Siskiyou Mountains Mt. Ashland site (13/14 individuals had this haplotype). Haplotype H was only observed in the Rocky Mountain and Alaskan sites. Haplotypes A and B are separated by a single nucleotide substitution, but haplotype H is separated by 2 nucleotide substitutions from haplotype B and 3 substitutions from haplotype A.

The average nucleotide diversity (π) across all sites was 1.88393 ± 2.16338 . Haplotype diversity ranged from 0, in fixed locations, to 0.8444 ± 0.1029 at Upper Slide Lake (Table 1). Pairwise Φ_{ST} values among populations ranged from 0.00 (between fixed sites in the Sierra Nevada) to 0.94 in pairwise comparisons of Sierran sites and Mt. Ashland (Table 2). Pairwise comparisons between Lake Earl and all other sites yielded Φ_{ST} values greater than 0.4, indicating moderate differentiation from all other sites. Although both Tajima's D and Fu's F values were generally negative, most values were not significantly different than zero (Table 3), supporting a neutral drift-mutation model for this gene, without indication of recent population expansion. However, it should be noted that two locations had significantly negative Tajima's D values, Bull Creek (Tajima's D = -1.79616, $p=0.029$) and Leek Springs (Tajima's D = -2.2777, $p=0.001$). Carson Pass was the only population to have a negative, significant Fu's F (-2.19999, $p=0.015$). A mismatch distribution analysis supports the idea of neutrality with low and non-significant values of the raggedness index, r .

Regions accounted for a substantial partitioning of molecular variance, with much of the variance also observed within populations in all AMOVA analyses. The largest amount of genetic partitioning explained by region is accounted for by the SAMOVA

k=8 (Figure 4, Table 4) clustering (Φ_{CT} 0.43936, Φ_{SC} of 0.02066, and Φ_{ST} of 0.45094), while the taxonomic (Φ_{CT} =0.27040, Φ_{SC} =0.19761, and Φ_{ST} = 0.41457) and *a priori* (Φ_{CT} of 0.31389, Φ_{SC} of 0.17631, and Φ_{ST} of 0.43486) designated regions (Table 5). The SAMOVA analysis clustered all Rocky Mountain populations together, all Sierra sites plus the nearby Sweetwater Mountains site and the Plaskett Meadows site in the Coast Range created another large group and all other sites (Mt. Ashland, White Mountains, Lake Earl, Cave Lake, Dismal Swamp, and Alaska) made up their own unique groups. This result differs from the *a priori* regional grouping based on mountain ranges by splitting the geographically proximal Warner mountain locations (Cave Lake and Dismal Swamp), as well as those in the Coast range, while adding the Sweetwater Mountains and Plaskett Meadows to the Sierran group (Figure 5). The NMDS analysis of pairwise Φ_{ST} values also conforms to the k=8 clustering, with similar groupings, Plaskett Meadows and the Sweetwater Mountains clustering near all Sierra Nevada populations, and the Rocky Mountains forming their own group, and all other locations falling into their own groups (Figure 6).

Because some geographic clustering of sample sites (clustering of all Sierra Nevada sites and clustering of all Rocky Mountain sites) was observed in the SAMOVA and NMDS analyses of pairwise Φ_{ST} values a Mantel's test was used to test for isolation by distance across all sites. The Mantel's test for isolation by distance across all sample sites was not significant ($r=0.0349$, $p=0.3896$), however, this pattern is likely an artifact of the sampling sites (Figure 7). There is a substantial area of unsampled terrain separating the Alaskan site from all others. A second Mantel's test revealed increasing

genetic differentiation with geographic distance is occurring throughout the contiguous United States ($r=0.445$ $p=0.0057$)(Figure 7).

ii. Morphological Analyses

163 individuals were included in morphological analysis of wing pigment patterns in *P. saepiolus*, including 7-13 males from each of the 16 sample locations used in the mtDNA analysis. All individuals used in the genetic analyses were evaluated for measurement; individuals with worn or missing wings were omitted from morphological measurements. Additional individuals from each population, not used in the genetic analyses, were also evaluated based on wing condition in order to increase sample size. A total of 19 wing spots on the right ventral hind wing were measured for spot area and corrected for wing size variation by dividing each spot area by total wing area. Principal components analysis of spot area showed 45.94 % of total variance and 13.42 % of the total variance explained by principal components 1 and 2 respectively (Figure 8). Although, a large amount of the variance was explained by the first principal component, here is little pattern evident that corresponds to distinct populations or the groupings based on mtDNA as identified by the SAMOVA, wing pigment spot pattern variation does not appear to conform to genetic patterns.

Multiple analyses of variance were conducted to determine if significant differences in wing spot pattern were present across populations and genetically defined regions using scores of the first two principal components (Table 6). In all analyses, significant differences among groups were detected ($p=0.0012$ and $p=0.00$). The Tukey's HSD analyses found significant differences between mean principal component scores only in comparisons of the Lake Earl location with other groups in both the population

and genetic clustering-based analyses (Table 7). This seems to be driven mainly by overall diminishment or absence of some wing spots in the Lake Earl population. The Lake Earl population, however, did not significantly differ from the other northern California and Oregon border populations (Cave Lake, Dismal Swamp, and Mt. Ashland), nor did it differ significantly from the Alaska population.

IV. Discussion

The mitochondrial sequence variation in *P. saepiolus* falls out into 8 distinct groups, largely concordant with mountain ranges. However, there are some individual populations that are exceptions to this, including the single population groupings in Northern California of Lake Earl, Dismal Swamp and Cave Lake. The AMOVA and SAMOVA of this grouping explain a substantial amount of the variance. The large amount of closely related haplotypes throughout the population suggests recent radiation within the species. Additionally, this species exhibits genetic signatures consistent with mainly the Alpine archipelago refuge model with southwestern patterns of dispersal across the landscape. The wing pattern in *P. saepiolus*, with small exception, does not correspond to mtDNA variation. The Lake Earl site of the coast range is one group that is distinct from several others, mirroring the genetic distinction of this population. Because there is little agreement in the genetic and wing pattern variation across populations the validity of current subspecific taxonomy must be re-evaluated.

i. Genetic variation within populations of *P. saepiolus*

Like other Lepidoptera species in the western United States, *P. saepiolus*, exhibits signs of recent population expansion with several rare alleles separated by few nucleotide

substitutions (Schoville et al. 2011). Based on the haplotype network there is little population structuring, with 2 widespread common haplotypes and several haplotypes unique to a single site. According to the AMOVA analyses, regardless of regional grouping Φ_{ST} , the variance explained by variation within populations was greater than Φ_{CT} , the variance explained by the variation across regions. This pattern is likely due to the large number of unique haplotypes in populations like Fairbanks (4 unique haplotypes), Leek Springs (3 unique haplotypes), Carson Pass (3 unique haplotypes), and Upper Slide Lake (3 unique haplotypes). The high overall haplotype diversity across the range is unusual as other studies of mitochondrial variation in North American invertebrates generally have fewer haplotypes (Nice and Shapiro 2001, Schoville et al. 2011).

ii. Genetic structuring of *P. saepiolus* by region and biogeographical hypotheses

Plebejus saepiolus does not appear to have deep structuring within mountain ranges as is seen in other Lepidoptera, and altitudinal changes across ranges, like the Sierra- Nevada and Rocky Mountains do not appear to inhibit migration between sites as it does in *Colias behrii* (Schoville et al. 2011). Though regional grouping generally explained less variance than that within populations, Φ_{CT} values were relatively large and there does appear to be geographic structuring of genetic variance across the contiguous United States. This is supported by a moderate and significant pattern of isolation by distance when not considering the Fairbanks population.

Genetic structuring in *P. saepiolus* appears complex, and can be characterized by different expectations of several biogeographical models. The regional distribution of COII haplotypes in *P. saepiolus* exhibits some expectations of the expanding Alpine

Archipelago model, with shared haplotypes across regions, however Tajima's D , Fu's F and Harpending's Raggedness index do not support the expectation of rapid population expansion during glacial periods (Knowles 2001, DeChaine and Martin 2004, Schoville and Roderick 2009, Schoville et al. 2011). Rather, shared haplotypes may be indicative of the ancestral polymorphism. As there is a moderate pattern of isolation by distance it is possible that several small dispersal events occurred through the western United States during glacial periods with colonization occurring during interglacial periods as is consistent with the alpine archipelago refuge model. It is possible that unique haplotypes in Alaska (UAF) represent ancestral haplotypes that did not reach the contiguous United States, acting as an isolated refuge following initial colonization of the southern part of the species range. Also consistent with this model is the large amount of variation shared between regions, with only a single, uncommon haplotype (haplotype I) between the Rocky Mountains and the Sierra Nevada. It seems likely that colonization occurred from the North to the Rocky Mountains, as the Alaska population only shares a haplotype with the Rocky Mountain population (haplotype H), aside from one individual with the widespread haplotype B. Those haplotypes most closely related to haplotype I and other Rocky Mountain and Alaska haplotypes are primarily observed in Northern California populations (Dismal Swamp, Cave Lake and Mt. Ashland). These sites may have been colonized separately from the Sierra Nevada via Alaska or the Rocky Mountains.

If colonization occurred leptokurtically, as there is some genetic clustering by region, it seems likely that sites with lower haplotype diversity, like Lake Earl, Mount Ashland and the White Mountains, were colonized after the Sierra Nevada, Rocky Mountains and Alaska populations which have higher observed haplotype diversity

(Cann et al. 1987, Templeton 1998, Nice and Shapiro 2001, Petit 2011). This pattern indicates dispersal routes occurring from the North and East toward western populations.

iii. Variation in wing pattern and subspecific designations within *P. saepiolus*

Variation in hind wing spot pattern does not agree with COII mitochondrial structuring across the landscape. There is no apparent pattern of variation in the biplot illustrating variation explained across principal components 1 and 2. The significant differences in wing spot variation in populations and genetically defined regions are primarily driven by the low-elevation Lake Earl population. The Lake Earl population occurs in a coastal region, and therefore, may have different habitat conditions than the others sampled, it is possible that the reduction in wing spot area is correlated with habitat differences.

Current subspecific designations do recognize the Lake Earl population as a unique site and this should be retained, as it is both genetically and morphologically distinct from other populations. As the morphological trait measured, hind wing spot area, did not differentiate between the other sites, additional subspecific grouping should be based on standing genetic variation.

The current designations of *P. s. littoralis* (Lake Earl), *P.s. albomontanus* (White Mountains), *P.s. amica* (Alaska), and *P. s. whitmeri* (Rocky Mountains) should be retained according to the genetic analyses. However, the Sierra Nevada group, which was previously split into Northern (*P.s. saepiolus*) and Southern (*P.s. aeheja*) regions can be collapsed, as well as incorporating the southern coast range populations (Plaskett Meadows), which was designated as *P.s. insulanus*. Additionally, the 3 other populations within *P. s. insulanus* should be given unique designations specifically for the Siskyou range, Dismal Swamp and Cave Lake populations. The separation of the

Cave Lake and Dismal Swamp populations merit further study, as the populations are in such close proximity to each other. It is possible that not all haplotypes were sampled from these regions, and genetic differentiation between the sites is inflated. Overall, the genetic and morphological variation observed in *P. saepiolus* suggests a complex biogeographic history and requires additional examination of life history traits of populations and additional molecular markers to further understand the demographic history and distribution of the species.

Table 1. Sampling Locations and Designations for *Plebejus saepiolus*.

Population	Nominal taxonomic designation	Mountain Range	N	COII haplotypes (#)	<i>h</i>	Group by Region	# of male wings
Carson Pass (CP)	<i>P.s. saepiolus</i>	Sierra Nevada	15	A(10), B(2), E(1), S(1), U(1)	0.562 ± 0.143	Sierra Nevada	10
Donner Pass (DP)	<i>P.s. saepiolus</i>	Sierra Nevada	14	A(14)	0	Sierra Nevada	8
Leek Springs (LS)	<i>P.s. saepiolus</i>	Sierra Nevada	15	A(11), B(1), K(1), T(1), AF(1)	0.476 ± 0.155	Sierra Nevada	12
Mount Rose (MR)	<i>P.s. saepiolus</i>	Sierra Nevada	16	A(16)	0	Sierra Nevada	11
Cave Lake (CL)	<i>P.s. insulanus</i>	Warner Mountains	13	A(3), B(6), V(1), AB(1), AE(2)	0.756 ± 0.097	Cave Lake	10
Dismal Swamp (DS)	<i>P.s. insulanus</i>	Warner Mountains	12	A(2), B(2), D(1), X(1), AB(1), AE(5)	0.818 ± 0.096	Dismal Swamp	11
Mount Ashland (MA)	<i>P.s. insulanus</i>	Siskyou Range	14	B(13), C(1)	0.143 ± 0.119	Siskyou	10
Plaskett Meadows (PL)	<i>P.s. insulanus</i>	Coast Range	15	A(7), F(2), G(6)	0.648 ± 0.0716	Sierra Nevada	13
Lake Earl (LE)	<i>P.s. littoralis</i>	Coast Range	13	B(2), W(11)	0.282 ± 0.142	Lake Earl	9
Sweetwater Mountains (SW)	<i>P.s. aeheja</i>	Sweetwater Mountains	15	A(5), B(5), J(1), R(1), Y(1), AB(2)	0.800 ± 0.071	Sierra Nevada	8
Tioga Crest (TI)	<i>P.s. aeheja</i>	Sierra Nevada	13	A(13)	0	Sierra Nevada	11
Fairbanks, AK (UAF)	<i>P.s. amica</i>	Alaska	15	B(1), H(3), I(1), N(6), O(1), P(3)	0.800 ± 0.071	Alaska	10
White Mountains (WH)	<i>P.s. albomontanus</i>	White Mountains	12	A(2), Q(1), Z(9)	0.439 ± 0.158	Whites	11
Bull Creek (BC)	<i>P.s. whitmeri</i>	Rocky Mountains	14	D(2), H(11), AC(1)	0.385 ± 0.149	Rocky Mountains	10
Soda Lake (SL)	<i>P.s. whitmeri</i>	Rocky Mountains	15	D(5), H(9), J(1)	0.562 ± 0.954	Rocky Mountains	11
Upper Slide Lake (USL)	<i>P.s. whitmeri</i>	Rocky Mountains	10	D(2), H(4), L(1), M(1), AC(1), AD(1)	0.844 ± 0.103	Rocky Mountains	7
Total			221				162

Table 2. Pairwise Φ_{ST} Values for All Sampling Location Comparisons. Extremely differentiated populations are bold, and undifferentiated comparisons are filled grey.

	BC	SL	USL	CP	DO	LS	MR	TC	CL	DS	LE	PL	MA	SW	UAF	WH
BC	0															
SL	0	0														
USL	0.00237	0.08417	0													
CP	0.63871	0.72151	0.40377	0												
DO	0.73885	0.8471	0.4785	0.0276	0											
LS	0.405	0.43563	0.23606	0	0	0										
MR	0.7535	0.85632	0.50329	0.03852	0	0.0053	0									
TC	0.73083	0.84201	0.46501	0.02155	0	0	0	0								
CL	0.33275	0.37065	0.15276	0.15244	0.20135	0.0635	0.22039	0.19115	0							
DS	0.43107	0.47351	0.27369	0.40606	0.43563	0.28951	0.45622	0.42322	0.05992	0						
LE	0.72513	0.7988	0.54176	0.76659	0.90044	0.45622	0.90722	0.89666	0.38767	0.48251	0					
PL	0.67548	0.76053	0.43655	0.15589	0.27233	0.01598	0.29156	0.26204	0.17802	0.40035	0.79866	0				
MA	0.61788	0.71903	0.36757	0.61533	0.93347	0.21558	0.93797	0.93097	0.12963	0.37565	0.8045	0.71527	0			
SW	0.41919	0.4663	0.19952	0.07298	0.13777	0	0.15341	0.12935	0.01925	0.27044	0.49948	0.14005	0.17842	0		
UAF	0.29204	0.32337	0.16571	0.44591	0.57263	0.24703	0.59162	0.56235	0.16711	0.3584	0.5798	0.50421	0.22361	0.21428	0	
WH	0.59223	0.66708	0.37506	0.53466	0.71847	0.24253	0.73514	0.70928	0.18629	0.36823	0.46812	0.60153	0.49579	0.24649	0.35667	0

Table 3. Tajima's D and Fu's F Values for All Sampling Locations. Respective *P* values for each value listed. (Bold denotes values significantly different from zero).

Population	Tajima's D	<i>P</i> value	Fu's F	<i>P</i> value
Carson Pass	-1.19547	0.112	-2.19999	0.014
Donner Pass	0	1	0	N/A
Leek Springs	-2.2777	<0.001	0.45342	0.794
Mount Rose	0	1	0	N/A
Cave Lake	-0.42797	0.386	2.41038	0.869
Dismal Swamp	1.24767	0.922	2.25591	0.84
Mount Ashland	-1.15524	0.151	-0.59478	0.11
Plaskett Meadows	0.62806	0.754	0.3653	0.544
Lake Earl	-0.3504	0.341	1.31652	0.683
Sweetwater Mountains	-0.68591	0.28	0.05629	0.518
Tioga Crest	0	1	0	N/A
Fairbanks, Alaska (UAF)	0.76339	0.784	-0.97962	0.251
White Mountains (McAfee Meadow)	-0.74109	0.25	0.77673	0.651
Bull Creek	-1.79616	0.015	1.80276	0.833
Soda Lake	-0.45679	0.348	0.41436	0.55
Upper Slide Lake	-0.11553	0.456	-0.05856	0.441

Table 4. SAMOVA Results. Optimal Φ_{CT} values for k=2-15 clustering.

k	Φ_{CT}
2	0.35933
3	0.38308
4	0.42125
5	0.42775
6	0.43893
7	0.44722
8	0.45265
9	0.4545
10	0.45841
11	0.45744
12	0.45654
13	0.45867
14	0.45878
15	0.45266

Table 5. Results of all AMOVA Analyses. Hierarchical analysis of molecular variance in COII mtDNA sequence data (A) Results grouping populations by mountain ranges (*a priori* grouping) (B) Results grouping populations by current subspecies designations (C) Results grouping populations by regions designated by SAMOVA clustering.

A. AMOVA by *a priori* grouping

Source of Variation	d.f.	SSD	Variance Component	% of total	<i>P</i> value
Among Ranges	7	112.921	0.49116	31.39	<0.001
Among Populations	8	27.968	0.18929	12.1	0.003
Within Populations	205	181.287	0.88433	56.51	<0.001
Totals	220	322.176	1.56479		

B. AMOVA by taxonomic designation

Source of Variation	d.f.	SSD	Variance Component	% of total	<i>P</i> value
Among Subspecies	6	0.231	0.00098	27.04	<0.001
Among Populations	9	0.084	0.00052	14.42	<0.001
Within Populations	205	0.435	0.00212	58.54	<0.001
Totals	220	0.57	0.00362		

C. AMOVA by SAMOVA/NMDS clustering

Source of Variation	d.f.	SSD	Variance Component	% of total	<i>P</i> value
Among Ranges	7	0.293	0.0017	43.94	<0.001
Among Populations	8	0.022	0.00004	1.16	0.01
Within Populations	205	0.435	0.00212	54.91	<0.001
Totals	220	0.75	0.00386		

Table 6. Results of One-Way ANOVAs for Principle Component 1. Using from wing area analysis (A) Results grouping by population (B) Results grouping by SAMOVA defined regions.

A.

ANOVA of Principal Component 1 by Population

Source of Variation	d.f.	F	<i>P</i> value
Population	15	2.229	0.00772
Residuals	147		

B.

ANOVA of Principal Component 1 by Genetic Cluster

Source of Variation	d.f.	F	<i>P</i> value
Clusters	7	3.651	0.00112
Residuals	155		

Table 7. Tukey's HSD Comparisons by Sampling Locations and Groups. (A) Population pairwise comparisons significantly different than zero with reported *P*-value, estimate of population-wise difference in PC1 with confidence interval (B) All pairwise comparisons between SAMOVA defined regions (Bold indicates values significantly different from zero).

A.

Populations	<i>P</i> value	Estimate (CI)
Lake Earl-Bull Creek	0.01668	-5.51800 (-10.53750, -0.49843)
Lake Earl-Leek Springs	0.02132	5.19360 (0.37628, 10.01093)
Lake Earl-Sweetwater Mountains	0.00863	5.94300 (0.79307, 11.09294)
Lake Earl-Upper Slide Lake	0.03735	5.65761 (0.15210, 11.16313)
Lake Earl-White Mountains	0.02211	5.27824 (0.36797, 10.18851)

B.

	Rockies	Cave Lake	Sierra Nevada	Dismal Swamp	Lake Earl	Mount Ashland	Fairbanks, AK
Rockies							
Cave Lake	0.92241						
Sierra-Nevada	0.93512	0.99924					
Dismal Swamp	0.99510	0.99989	1.00000				
Lake Earl	0.00144	0.21740	0.00599	0.07308			
Mount Ashland	0.98268	1.00000	1.00000	1.00000	0.12850		
Fairbanks, AK	0.06866	0.84502	0.22432	0.56717	0.95974	0.70704	
White Mountains	0.99993	0.91133	0.95161	0.98901	0.00598	0.97226	0.12472

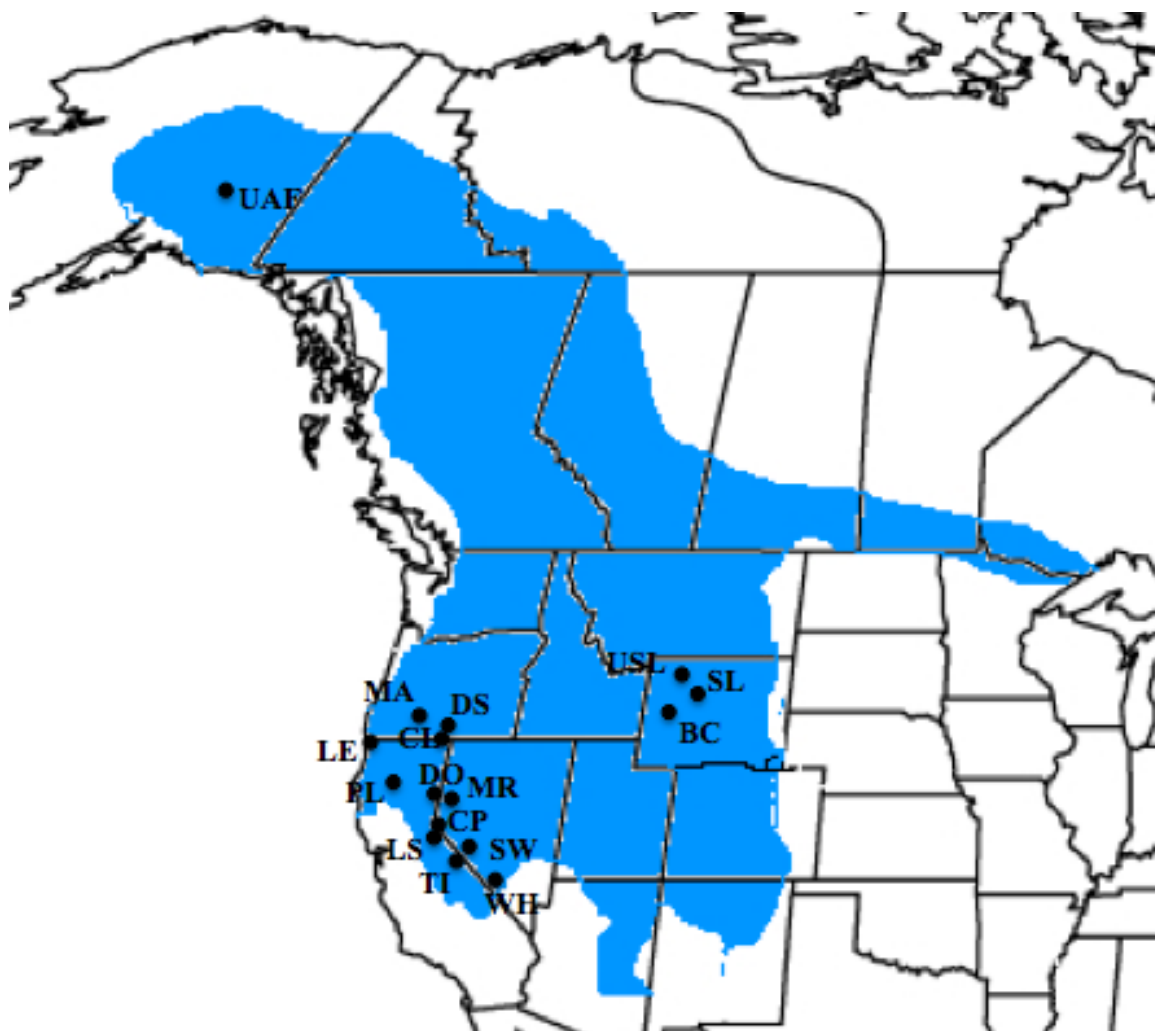
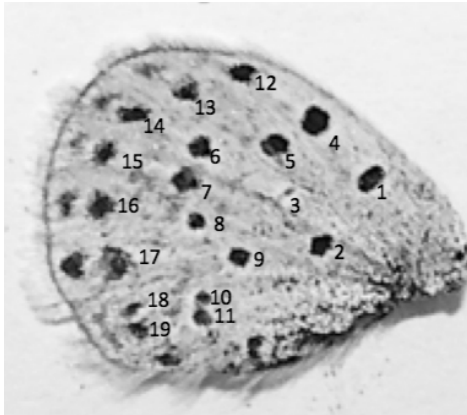


Figure 1. Map of Sampling Locations and Range Map for *Plebejus saepiolus*.

A.



B.

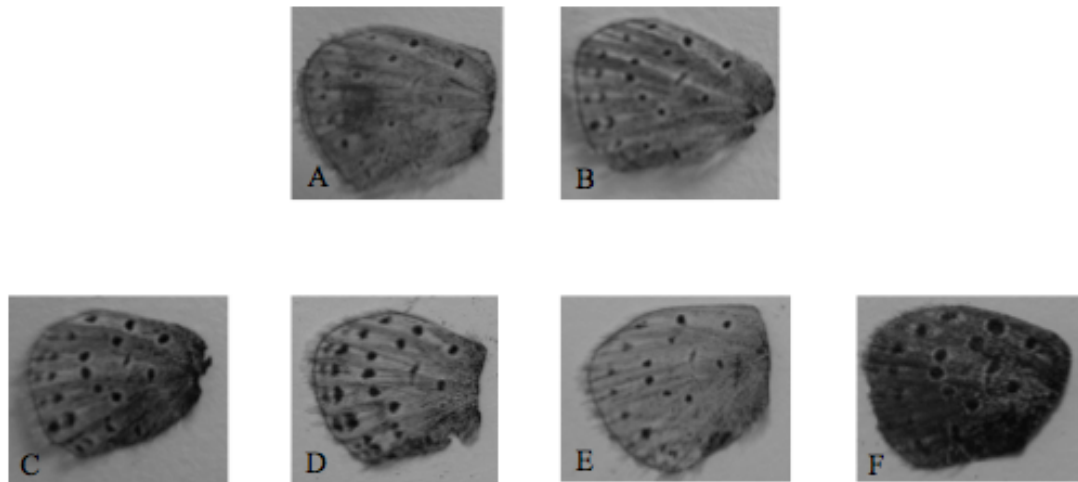
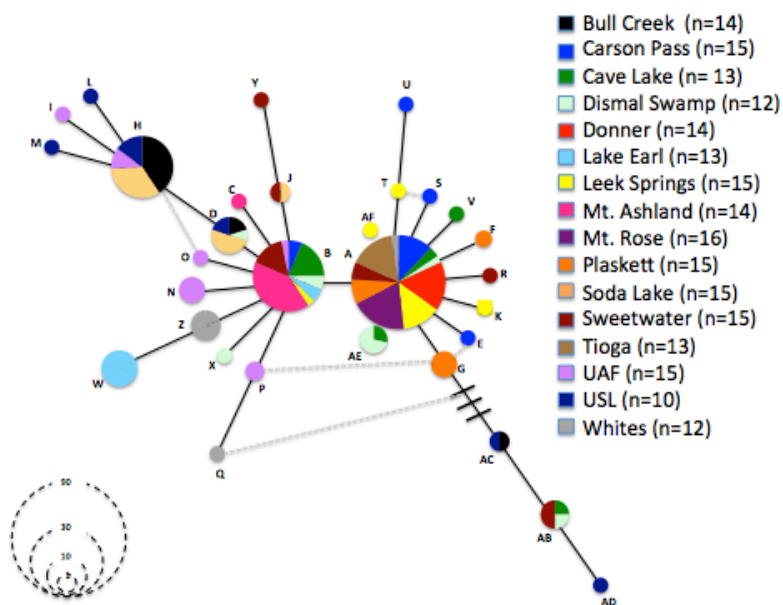


Figure 2. Wing Spot Measurements and Examples from Differentiated Populations. A. Example of hind wing spot areas measured. Nineteen spots measure for area. (Marginal spot row and inferior spots were not included in analysis because of variable wing condition). B. Examples of wing spot pattern in Lake Earl (A) and Fairbanks, AK (B) regions compared to the Rocky Mountains (Individual from Bull Creek (C)), the Sierra Nevada (Individual from Leek Springs (D)), Mt. Ashland (E) and White Mountains (F) regions.

A.



B.

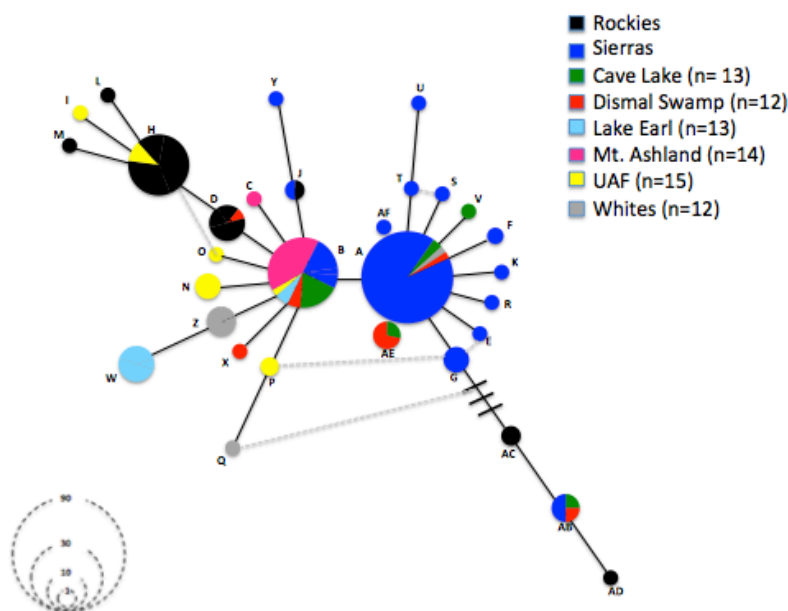


Figure 3. Haplotype Networks by Population and Groups. Circles indicates the number of individuals per haplotype. Dashed lines represent equally parsimonious connections. Hashes represent single nucleotide changes in unrepresented haplotypes. A. Haplotypes divided by population B. Haplotypes divided by SAMOVA clustering.

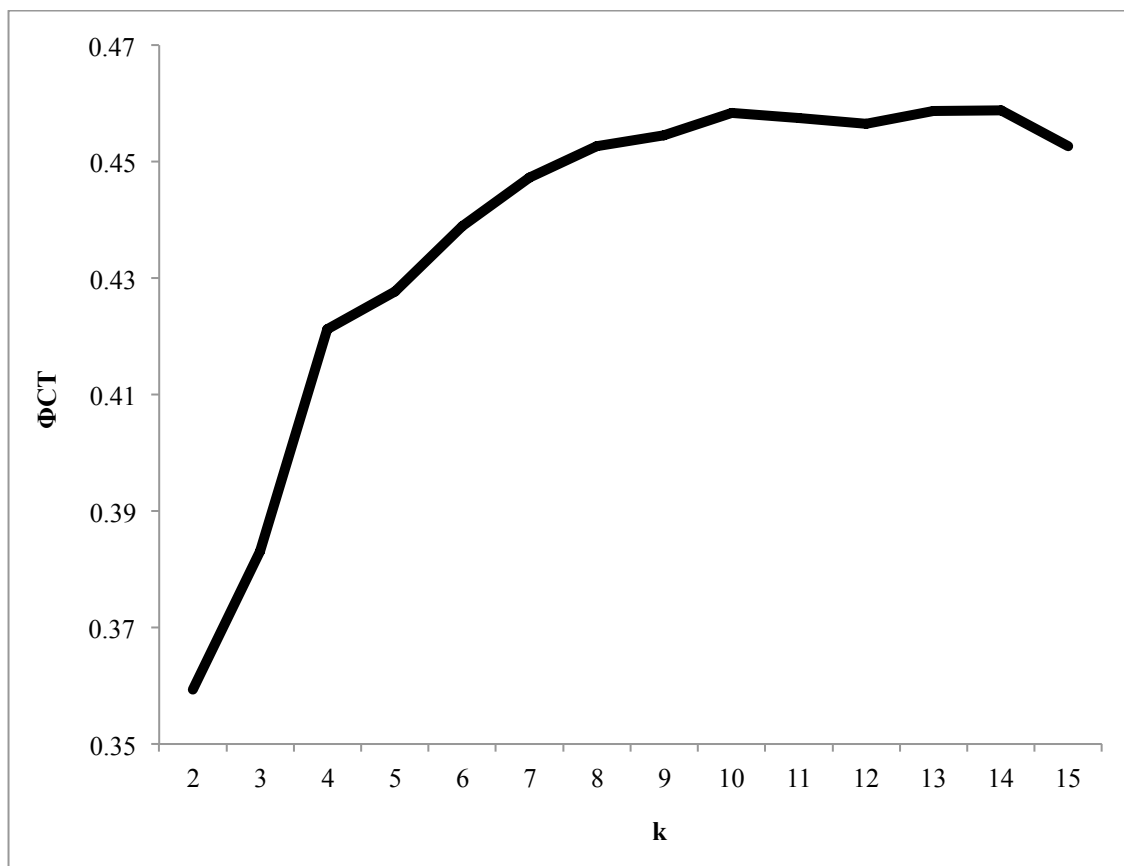


Figure 4. Graph Showing Optimized Φ_{ST} by Number of Groups (k).

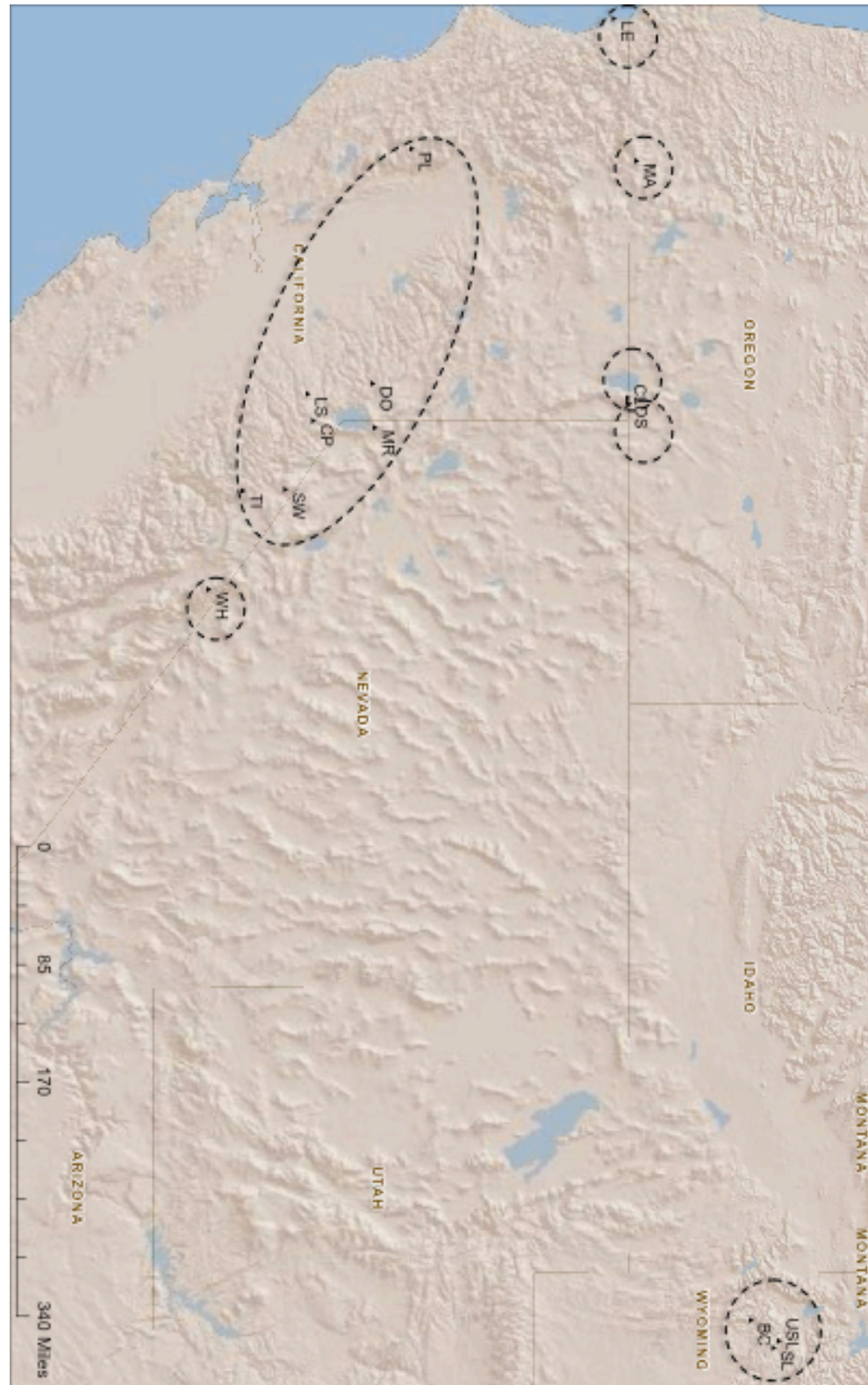


Figure 5. Map of Population Groupings Based on SAMOVA Results. (Alaska population not shown).

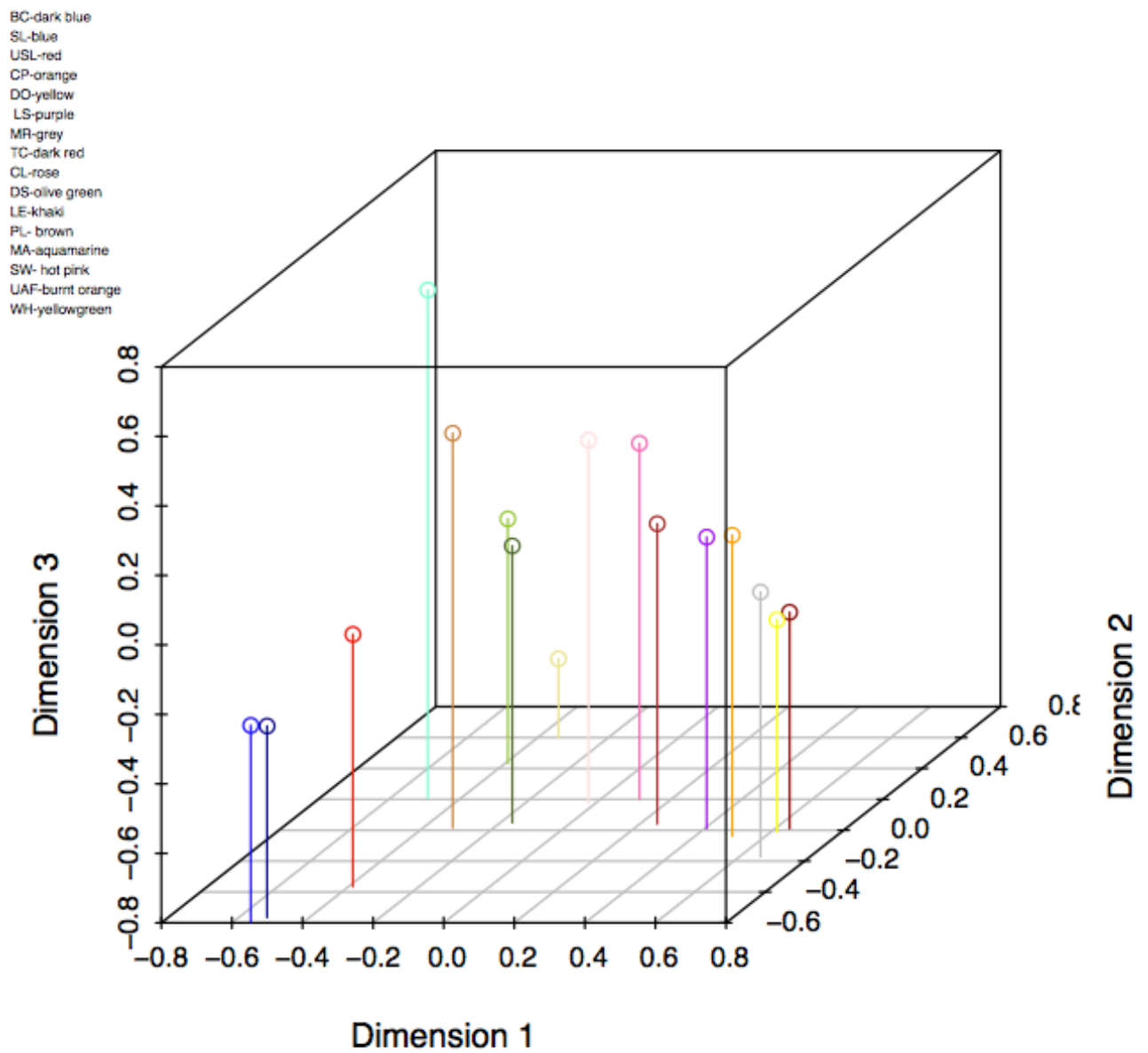
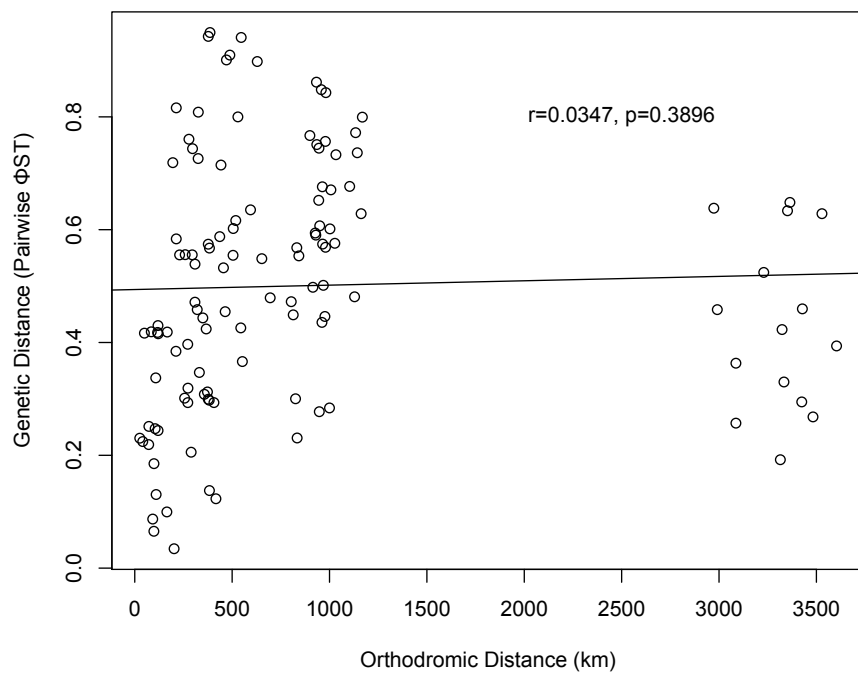


Figure 6. Non-metric Multi Dimensional Scaling (NMDS) plot of Pairwise F_{ST} 's for All Populations.

A.



B.

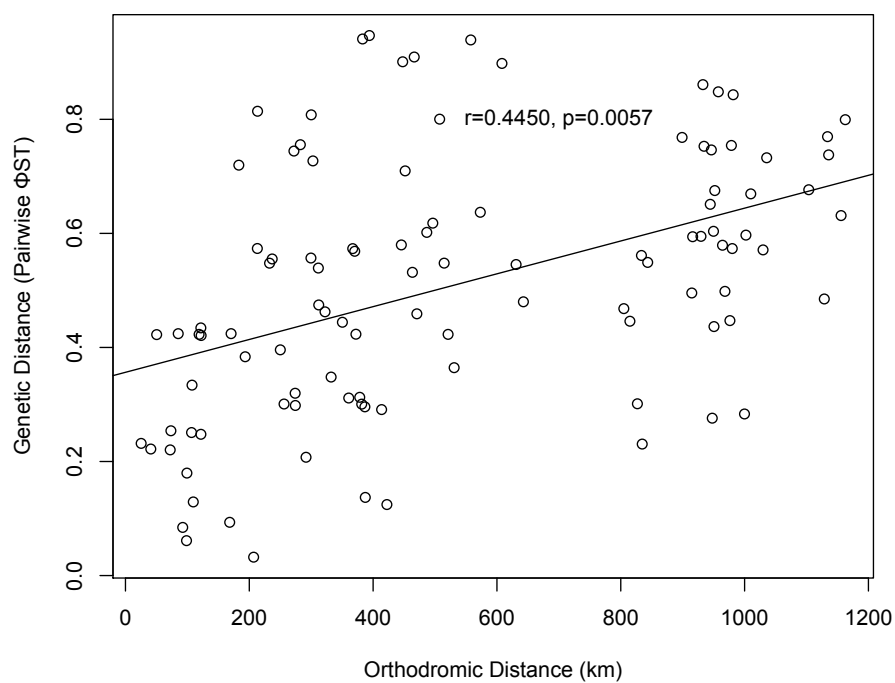


Figure 7. Graphs of Genetic Distance by Geographic (Orthodromic) Distance. r and p -values of isolation by distance analysis (A) All populations (B) Excluding Fairbanks, AK population.

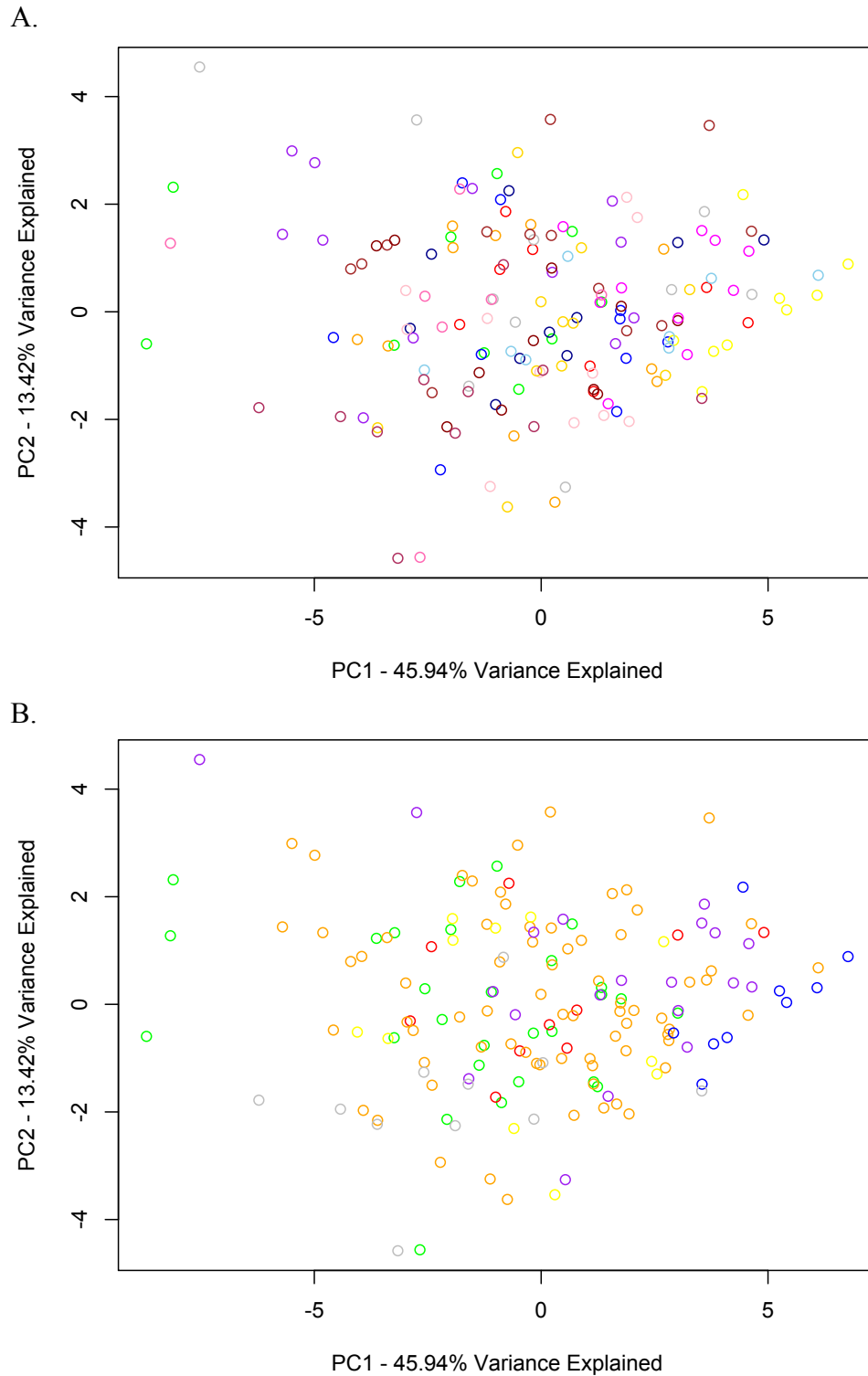


Figure 8. Principle Components Analysis (PCA) Biplots by Populations and Groups. Principal Components Scores 1 and 2 of wing spot area (A) Colored by Population and (B) Colored by SAMOVA regional grouping.

APPENDIX

Appendix 1. Unique haplotype sequences of 492bp region of COII mtDNA

Haplotype	Sequence
A	TATTAATCGATTCTTATTAGAAGGTCAAATAATTGAATTAATTTGAACTATTTTACCAGCTATTACTTTA ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACCA TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCCGATTTCAAAAATATTGAATTTGATT CATATATAATAAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC CATAAATAATAATATTCGTATATTAATTACTGCAACTGATGTAATTCACTCATGAACTGTCCCATCTAT TGGTGTTAAAGTTGATGCTAATCCTGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGGAAT TTTTTTTGGTCAATGTTTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAAT TCAAT
B	TATTAATCGATTCTTATTAGAAGGTCAAATAATTGAATTAATTTGAACTATTTTACCAGCTATTACTTTA ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACCA TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCCGATTTCAAAAATATTGAATTTGATT CATATATAATAAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC TATAAATAATAATATTCGTATATTAATTACTGCAACTGATGTAATTCACTCATGAACTGTCCCATCTATT GGTGTTAAAGTTGATGCTAATCCTGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGGAAT TTTTTTTGGTCAATGTTTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAAT TCAAT
C	TATTAATCGATTCTTATTAGAAGGTCAAATAATTGAATTAATTTGAACTATTTTACCAGCTATTACTTTA ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACCG TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCCGATTTCAAAAATATTGAATTTGATT CATATATAATAAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC TATAAATAATAATATTCGTATATTAATTACTGCAACTGATGTAATTCACTCATGAACTGTCCCATCTATT GGTGTTAAAGTTGATGCTAATCCTGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGGAAT TTTTTTTGGTCAATGTTTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAAT TCAAT
D	TATTAATCGATTCTTATTAGAAGGTCAAATAATTGAATTAATTTGAACTATTTTACCAGCTATTACTTTA ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACCA TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCCGATTTCAAAAATATTGAATTTGATT CATATATAATAAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC TATAAATAATAATATTCGTATATTAATTACTGCAACTGATGTAATTCACTCATGAACTGTCCCATCTATT GGTGTTAAAGTTGATGCTAATCCCGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGGAAT TTTTTTTGGTCAATGTTTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAAT TCAAT
E	TATTAATCGATTCTTATTAGAAGGTCAAATAATTGAATTAATTTGAACTATTTTACCAGCTATTACTTTA ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACCA TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCCGATTTCAAAAATATTGAATTTGATT CATATATAATAAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC CATAAATAATAATATTCGTATATTAATTACTGCAACTGATGTAATTCACTCATGAACTGTCCCATCTAT TGGTGTTAAAGTTGATGCTAATCCTGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCTGGAAT TTTTTTTGGTCAATGTTTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAAT TCAAT

F	TATTAATCGATTCTTATTAGAAGGTCAAATAATTGAATTAATTTGAACTATTTTACCAGCTATTACTTTA ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACCA TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCTGATTTCAAAAAATATTGAATTTGATT CATATATAATAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC CATAAATAATAAATATTCGTATATTAATTACTGCAACTGATGTAATTCACCTCATGAACTGTCCCATCTAT TGGTGTTAAAGTTGATGCTAATCCTGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGGAAT TTTTTTTGGTCAATGTTTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAAT TTCAAT
G	TATTAATCGATTCTTATTAGAAGGTCAAATAATTGAATTAATTTGAACTATTTTACCAGCTATTACTTTA ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACCA TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCCGATTTCAAAAAATATTGAATTTGATT CATATATAATAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC CATAAATAATAAATATTCGTATATTAATTACTGCAACTGATGTAATTCACCTCATGAACTGTCCCATCTAT TGGTGTTAAAGTTGATGCTAATCCTGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGGAAT TTTTTTTGGTCAATGTTTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAAT TTCAAT
H	TATTAATCGATTCTTATTAGAAGGTCAAGATAATTGAATTAATTTGAACTATTTTACCAGCTATTACTTTA ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACCA TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCCGATTTCAAAAAATATTGAATTTGATT CATATATAATAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC TATAAATAATAAATATTCGTATATTAATTACTGCAACTGATGTAATTCACCTCATGAACTGTCCCATCTATT GGTGTTAAAGTTGATGCTAATCCCGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGGAATT TTTTTTTGGTCAATGTTTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAATT TCAAT
I	TATTAATCGATTCTTATTAGAAGGTCAAGATAATTGAATTAATTTGAACTATTTTACCAGCTATTACTTTA ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACCA TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCCGATTTCAAAAAATATTGAATTTGATT CATATATAATAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC TATAAATAATAAATATTCGTATATTAATTACTGCAACTGATGTAATTCACCTCATGAACTGTCCCATCTATT GGTGTTAAAGTTGATGCTAATCCCGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGGAATT TTTTTTTGGTCAATGTTTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAATT TCAAT
J	TATTAATCGATTCTTATTAGAAGGTCAAATAATTGAATTAATTTGAACTATTTTACCAGCTATTACTTTA ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACCA TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCCGATTTCAAAAAATATTGAATTTGATT CATATATAATAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC TATAAATAATAAATATTCGTATATTAATTACTGCAACTGATGTAATTCACCTCATGAACTGTCCCATCTATT GGTGTTAAAGTTGATGCTAATCCTGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGGAATT TTTTTTTGGTCAATGTTTCAGAAATTTGTGGAGCAAATCATAGTTTTATACCTATTGTAATTGAAAGAATT TCAAT
K	TATTAATCGATTCTTATTAGAAGGTCAAATAATTGAATTAATTTGAACTATTTTACCAGCTATTACTTTA GTTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACCA TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCCGATTTCAAAAAATATTGAATTTGATT CATATATAATAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC CATAAATAATAAATATTCGTATATTAATTACTGCAACTGATGTAATTCACCTCATGAACTGTCCCATCTAT TGGTGTTAAAGTTGATGCTAATCCTGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGGAAT TTTTTTTGGTCAATGTTTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAAT TTCAAT

Y	TATTAATCGATTCTTATTAGAAAGGTCAAATAATTGAATTAATTTGAACTATTTTACCAGCTATTACTTTA ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACCA TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCCGATTTCAAAAAATATTGAATTTGATT CATATATAATAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC TATAAATAATAATATTCGTATATTAATTACCGCAACTGATGTAATTCACTCATGAACTGTCCCATCTAT TGGTGTAAAGTTGATGCTAATCCTGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGAAT TTTTTTGGTCAATGCTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAAT TTCAAT
Z	TATTAATCGATTTTATTAGAAAGGTCAAATAATTGAATTAATTTGAACTATCTTACCAGCTATTACTTTA ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACTA TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCCGATTTTAAAAATATTGAATTTGATT CATATATAATAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC CATAAATAATAATATTCGAATATTAATTACTGCAACTGATGTAATTCACTCATGAACTGTCCCATCTAT TGGTGTAAAGTTGATGCTAATCCTGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGAAT TTTTTTGGTCAATGCTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAAT TTCAAT
A B	TATTAATCGATTTTATTAGAAAGGTCAAATAATTGAATTAATTTGAACTATCTTACCAGCTATTACTTTA ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACCA TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCCGATTTTAAAAATATTGAATTTGATT CATATATAATAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC CATAAATAATAATATTCGAATATTAATTACTGCAACTGATGTAATTCACTCATGAACTGTCCCATCTAT TGGTGTAAAGTTGATGCTAATCCTGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGAAT TTTTTTGGTCAATGCTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAAT TTCAAT
A C	TATTAATCGATTTTATTAGAAAGGTCAAATAATTGAATTAATTTGAACTATCTTACCAGCTATTACTTTG ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACTA TTAAATCAATCGGACATCAATGATATTGAAGATATGAATATTCCGATTTTAAAAATATTGAATTTGATT CATATATAATAAATGAATTTAATAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC CATAAATAATAATATTCGAATATTAATTACTGCAACTGATGTAATTCACTCATGAACTGTCCCATCTAT TGGTGTAAAGTTGATGCTAATCCTGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGAAT TTTTTTGGTCAATGCTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAAT TTCAAT
A D	TATTAATCGATTTTATTAGAAAGGTCAAATAATCGAATTAATTTGAACTATTTTGCCAGCTATTACTTTA ATTTTCATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAACTTAATAATCCATTAATTACCA TTAAATCAATTGGACATCAATGATATTGAAGATATGAATATTCCGATTTTAAAAATATTGAATTTGATT CATATATAATAAATGAATTTAACAATAATAATTTTCGTTTATTAGATGTTGATAATCGTATTATTATCCC TATAAATAATAACATTCGTATATTAATTACTGCAACTGATGTAATTCACTCATGAACTATCCCATCTATT GGTGTAAAGTTGATGCTAATCCTGGTCGGTAAATCAAACAAGATTTTTCATTAATCGACCGGAATT TTTTTTGGTCAATGCTCAGAAATTTGTGGAGCAAACCATAGTTTTATACCTATTGTAATTGAAAGAATT TCAAT
A E	TATTAACCGATTTTATTAGAAAGGTCAAATAATTGAATTAATTTGAACTATTTTACCAGCTATTACTTTA ATTTTTATTGCTTTACCTTCATTACGTTTATTATATTTACTTGATGAAATTAATAATCCATTAATTACTA TTAAATCAATTGGTCATCAATGATACTGAAGATATGAGTATTCTGATTTTAAAAATATCGAATTTGATT CATATATAATAAATGAATTTGATAATAATAATTTTCGTTTATTAGATGTTGATAATCGAATTATTATCCC TATAAATAATAATATTCGTATGTTAATTACTGCAACTGATGTGATTCACTCATGAACTATTCCATCTATT GGTGTAAAGTTGATGCTAATCCTGGTCGATTAAATCAAACAAGATTTTTCATTAATCGACCGGAATT TTTTTCGGTCAATGTTTCAGAAATTTGTGGAGCAAATCATAGTTTTATACCTATTGTAATTGAAAGAATT TCAAT

LITERATURE CITED

- Austin, G.T. and D.D. Murphy. 1987. Zoogeography of Great Basin butterflies: Patterns of distribution and differentiation. *Great Basin Naturalist* 47(2): 186-201.
- Avise, J.C., C. Giblin-Davidson, J. Laerm, J.C. Patton, and R.A. Lansman. 1979. Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. *Proceedings of the National Academy of Sciences* 76: 6694- 6698.
- Avise, J.C. 1994. *Molecular Markers, Natural History and Evolution*. Chapman and Hall, New York, NY, USA.
- Avise, J.C. 2000. *Phylogeography: The history and formation of species*. Harvard University Press, Cambridge, Massachussets.
- Billings, W.D. 1973. Alpine Phytogeography across the Great Basin. *Great Basin Naturalist Memoirs* 2:105-116.
- Brock, J.P. and K. Kaufman. 2003. *Kaufman field guide to butterflies of North America*. Houghton Mifflin Company, New York, NY.
- Cann, R. L., M. Stoneking and A.C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* 325: 31–36.
- Clement, M., D. Posada, K. and Crandall. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9(10): 1657-1660.
- Darriba, D., G.L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9(8): 772.
- Dray, S. and A.B. Dufour. 2007. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* 22(4): 1-20.
- Dupanloup, I., S. Schneider, and L. Excoffier. 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* 11: 2571–2581.
- Excoffier, L. and H.E.L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564-567.

- Fordyce, J.A., C.C. Nice, M.L. Forsiter and A.M. Shapiro. 2002. The significance of wing pattern diversity in the Lycaenidae: mate discrimination by two recently diverged species. *Journal of Evolutionary Biology* 15: 871–879.
- Galbreath, K.E., D.J. Hafner, K.R. Zamudio and K. Agnew. 2010. Isolation and introgression in the intermountain west: contrasting gene genealogies reveal the complex biogeographic history of the American Pika (*Ochotona princeps*). *Journal of Biogeography* 37: 344–362.
- Geneious version 5.3.6 created by Biomatters. Available from <http://www.geneious.com/>.
- Gillespie, A.R. and P.H. Zehfuss. 2004. Glaciations of the Sierra Nevada, California, USA. *Quaternary Glaciations-Extent and Chronology, Part II*. Editors: Ehlers, J. and P.L. Gibbard. pp. 51–62, Elsevier, Amsterdam.
- Glassberg, J. 2001. *Butterflies through Binoculars: the West*. A field guide to butterflies of western North America. Oxford University Press, New York, NY.
- Hartl, D.L and A.G. Clark. 1997. *Principles of Population Genetics*, Third Edition. IE-Macmillan. United Kingdom.
- Heron, D. Garry Oak Invertebrates Recovery Implementation Group. 2007. Recovery strategy for the Greenish Blue *insulanus* subspecies (*Plebejus saepiolus insulanus*) in British Columbia. Prepared for the B.C. Ministry of Environment, Victoria, BC.
- Hewitt, G.M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of Linnean Society* 58: 247–276.
- Hewitt, G.M. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405: 907–913.
- Hewitt, G.M. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society London* 359: 183–195.
- Ibrahim, K.M., R.A. Nichols and G.M. Hewitt. 1996. Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* 77: 282–291.
- Jennings, S.A. and D.L. Elliott-Fisk. 1993. Packrat midden evidence of late quaternary vegetation change in the White Mountains, California-Nevada. *Quaternary Research* 39: 214–221.
- Knowles, L.L. 2001. Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Molecular Ecology* 10: 691–701.
- Lomolino, M.V., B.R. Riddle, R.J. Whitaker, and J. H. Brown. 2010. *Biogeography* (4th Edition). Sinauer Associates, Inc. Sunderland, Massachusetts.

- Morefield, J.D. 1992. Spatial and ecologic segregation of phytogeographic elements in the White Mountains of California and Nevada. *Journal of Biogeography* 19(1): 33-50.
- Murphy, D.D. 1990. A report on the California butterflies listed as candidates for endangered status by the United States Fish and Wildlife Service. Draft report for California Department of Fish and Game Contract No. C-1755.
- Nice, C.C. and A.M. Shapiro. 2001. Patterns of morphological, biochemical, and molecular evolution in the *Oeneis chryxus* complex (Lepidoptera: Satyridae): A test of historical biogeographical hypotheses. *Molecular Phylogenetics and Evolution* 20(1): 111-123.
- Nice, C.C., N. Anthony, G. Gelembiuk, D. Raterman and R. French-Constant. 2005. The history and geography of diversification within the butterfly genus *Lycæides* in North America. *Molecular Ecology* 14: 1741-1754.
- Opler, P.A. 1999. A field guide to western butterflies, 2nd edition. Houghton Mifflin Company. New York, NY.
- Paradis E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289-290.
- Petit, R.J. 2011. Early insights into the genetic consequences of range expansion. *Heredity* 106: 203-204.
- Pielou, E.C. 1992. After the Ice Age: The Return of Life to Glaciated North America. The University of Chicago Press. Chicago, IL.
- Pyle, R. M. 2002. The Butterflies of Cascadia, A Field Guide to All the Species of Washington, Oregon, and Surrounding Territories. Seattle Audubon Society, Seattle, Washington.
- R Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Runquist, E.B., M.L. Forister and A.M. Shapiro. 2012. Phylogeography at large spatial scales: incongruent patterns of population structure and demography of Pan-American butterflies associated with weedy habitats. *Journal of Biogeography* 39: 382-396.
- Schmitt, T. 2007. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology* 4: 1-13.
- Schoville, S.D. and G.K. Roderick. 2009. Alpine biogeography of Parnassian butterflies during Quaternary climate cycles in North America. *Molecular Ecology* 18: 3471-3485.

- Schoville, S.D. and G.K. Roderick. 2010. Evolutionary diversification of cryophilic *Grylloblatta* species (Grylloblattodea: Grylloblattidae) in alpine habitats of California. *BMC Evolutionary Biology* 10:163.
- Schoville, S.D., G.K. Roderick, and D.H. Kavanaugh. 2012. Testing the “Pleistocene species pump” in alpine habitats: lineage diversification of flightless ground beetles (Coleoptera: Carabidae: *Nebria*) in relation to altitudinal zonation. *Biological Journal of the Linnean Society* 107: 95-111.
- Schoville, S.D., M. Stuckey and G.K. Roderick. 2011. Pleistocene origin and population history of a neoendemic alpine butterfly. *Molecular Ecology* 20: 1233–1247.
- Scott, J.A. 1992. The butterflies of North America. Stanford University Press.
- Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* 87: 651– 701.
- Stephenson, J.R. and G.M. Calcarone. 1999. Southern California mountains and foothills assessment: habitat and species conservation issues. General Technical Report GTR-PSW-172. Albany, CA: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture.
- Swofford, D. 2002. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4. Sinauer Associates, Sunderland, MA.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- Templeton, A. R. 1998. Nested clade analyses of phylogeographic data: Testing hypotheses about gene flow and population history. *Molecular Ecology* 7: 381–397.
- Venables, W.N. and B.D. Ripley. 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York.
- 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 paleoecology of *Polyommatus* blue butterflies show Beringia was a climate-regulated gateway to the new world. *Proceedings of the Royal Society of London B* 278(1719): 2737-2747.

VITA

Desirae Margaret Weyland was born in Sachsenhausen, Frankfurt am Main, Germany on June 24, 1986 to Greg Weyland and Heidi Neddenriep-Weyland. After completing high school in Douglas County, Nevada, Desirae entered the University of California, Davis in 2004. She received dual degrees in 2008, graduating with a B.S. in Wildlife, Fish and Conservation Biology, and a B.A. in Spanish. Following her undergraduate career, Desirae spent two field seasons employed with the US Forest Service in the Lake Tahoe Basin as a Biological (Wildlife) Technician. In 2010, she entered the Graduate College at Texas State University-San Marcos in the Biology Department. Following completion of her Master's degree, Desirae plans to continue her education in a Ph.D. program at the University of Nevada, Reno.

Permanent Address:

870 Chambers Lane

Markleeville, CA 96120

This thesis was typed by Desirae M. Weyland.