Sequence evolution among divergent mitochondrial haplotypes within species of *Junonia* butterflies

by

Bonnie S. McCullagh

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

In partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Biological Sciences

University of Manitoba

Winnipeg

Abstract

The New World *Junonia* butterflies include well-studied model organisms yet their phylogeny has not been resolved by traditional *cox1* DNA barcodes. Sixteen *Junonia* mitochondrial genomes were sequenced using next generation MiSeq technology. *Junonia lemonias,* an Old World species, has mitochondrial genome features typical of other Ditrysian Lepidoptera, and synteny is maintained throughout the genus. Analysis of *Junonia* mitogenomes produced a robust phylogeny that was used in combination with biogeographic information to infer that *Junonia* crossed the Pacific Ocean to invade the New World on 3 separate occasions. *Junonia vestina,* a high elevation species from the Andes Mountains, shows high altitude adaptation in the mitochondrial protein coding loci *atp6, atp8, cox1, cob, nad1,* and *nad2,* with the strongest effects seen in *cox1* and *nad1*. There is some overlap between these genes with human loci that have disease associations with the same amino acid positions which could help elucidate the function of high elevation mutations in *J. vestina*.

Acknowledgements

First, I would like to thank my advisor Dr. Jeffrey Marcus for all of his help and support during my master's program. Every time that I have doubted my abilities, Jeff has told me that I can do this, and that I can get through whatever my difficulties were. I would not be even a third the scientist I am without his guidance and assistance.

I also would like to thank the members of my committee, Dr. Bruce Ford, Dr. Barbara Sharanowski, and Dr. Anne Worley, for all of their feedback and help throughout my graduate career. I would not have made it this far without all of their assistance!

I wish to thank all of the collectors who contributed butterfly specimens for this research project. I also would like to thank Dr. Deborah Tsuyuki, at the Manitoba Institute of Child Health, for all of her help with next generation sequencing.

The Marcus lab has received financial support from the NSERC Discovery Grant program, the Canada Research Chair program, and a Canada Foundation for Innovation Award. I benefited from a University of Manitoba Graduate Fellowship, and an International Graduate Student Scholarship. I also received a bursary through the Financial Aid and Awards General Award Fund in the summer of 2015.

The members of the Marcus lab over the years were always helpful and a source of fun. So, thank you to Roohollah Abbasi, Amber Gemmell, Ashley Haverstick, Melanie Lalonde, Jacob Miller, and Melissa Peters.

Thank you to Joanne Seiff, for keeping me grounded, and making sure I had a home cooked meal sometimes!

A huge thanks goes to my fellow graduate students in the department, for all the fun and camaraderie. I learned so much from all of you and I am so glad I got to know you. Special thanks to Pauline Catling, Chris Deduke, Michelle Ewacha, Tazarve Gharajehdaghipoor, Melanie Lalonde, Danielle Mocker, Kelsey O'Brien, Dylan Pond, Jen Sojka, Christa Szumski, Dawn Wood, and Miles Zhang, for helping me stay (relatively) sane throughout grad school.

Thank you to my parents, who paid my tuition, and for my plane tickets home when I had to escape to warmer weather in the winter. Thank you to all the rest of my family.

Thank you to all of my friends back in the US who did their best to support me from far away. Daiya Gholston, Randi Grounds, Jenny Kiefer, Liz L., Maggie Mahan, Sarah Randall, Elizabeth Riley, and B. Robertson: thank you. I owe you.

Debbie Chiang, Miranda Morsch, Lauren Saccone: my partners in puns, friendship, and media consumption. I am humbled to have you as friends, and our visits were greatly needed recharges.

Susie Connor: My practically related bestie. Thank you for watching all the bad movies with me, and letting me crash your vacation.

Dedication

To my mom, Ruth McCullagh, who has never wavered in her support of me, although she did refuse to visit in the winter.

Table of Contents

List of Tables

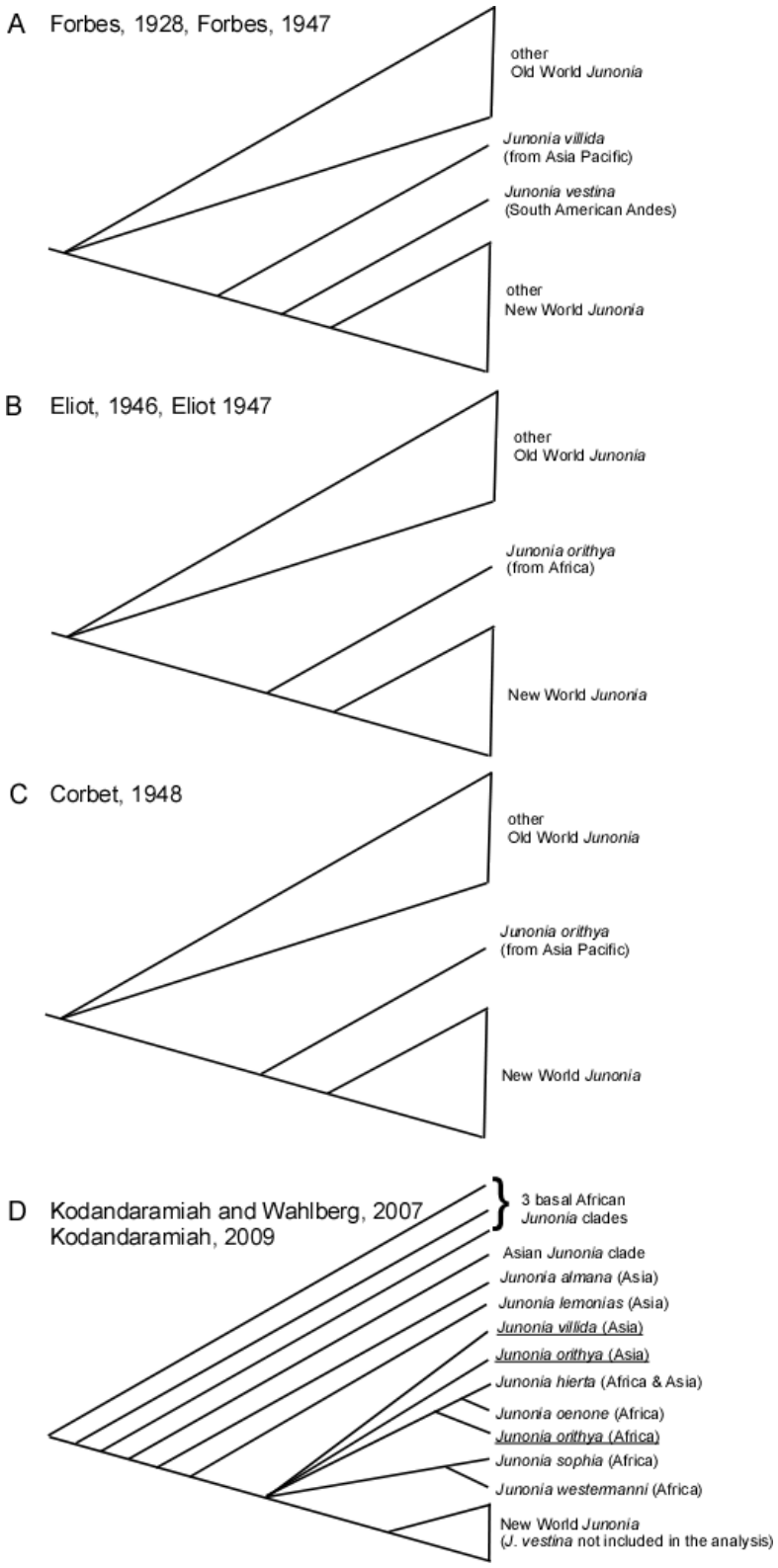
List of Figures

Chapter 1: Introduction

Diversity of butterfly genus Junonia

The genus *Junonia* (Lepidoptera: Nymphalidae: Nymphalinae) is a group of brush-footed butterflies that occurs in the Old and New Worlds. *Junonia* is found on every continent except Europe and Antarctica, and the New World species are found as far north as southern Canada and as far south as the Pampas of Argentina (Gemmell & Marcus, 2015). Although *Junonia* is a well-studied genus, *Junonia* taxonomy has been controversial. Understanding patterns of geographic, seasonal, intra-population, and interspecific variation has been challenging (Gemmell et al., 2014) and many *Junonia* species were formerly classified under the genus *Precis* (Wahlberg et al., 2005). However, in 2005, it was confirmed via molecular methods that *Junonia* and *Precis* are not synonymous, nor are they sister clades to each other, and most species were unambiguously assigned to the appropriate genus (Wahlberg et al., 2005). All of the New World species in this group were assigned to genus *Junonia*.

Fig. 1-1. Hypotheses regarding the origin of the New World *Junonia*. A, B, and C (Forbes, 1928, Eliot, 1946, Eliot, 1947, Forbes, 1947, Corbet, 1948) are morphology based phylogenetic hypotheses, which are primarily based on different aspects of *Junonia* wing color patterns. D is a molecular based phylogeny using partial sequences from mitochondrial *cox1*, and nuclear *ef1-alpha*, and *wingless* loci (Kodandaramaiah & Wahlberg, 2007, Kodandaramaiah, 2009). Species that play important roles in the morphology based phylogenetic hypotheses for the origin of the New World *Junonia* taxa are underlined in panel D. Only 3 of 10 New Worldspecies (*J. coenia*, *J. evarete*, and *J. genoveva*) were included in these molecular phylogenetic analyses and key species such as *J. vestina* were omitted.


A Forbes, 1928, Forbes, 1947
other
Old World Junonia
Junonia villida
(from Asia Pacific)
Junonia vestina
(South American Andes)
other
New World Junonia
B Eliot, 1946, Eliot 1947
other
Old World Junonia
Junonia orithya
(from Africa)
New World Junonia
C Corbet, 1948
other
Old World Junonia
Junonia orithya
(from Asia Pacific)
New World Junonia
D Kodandaramiah and Wahlberg, 2007
Kodandaramiah, 2009
3 basal African
Junonia clades
Asian Junonia clade
Junonia almana (Asia)
Junonia lemonias (Asia)
Junonia villida (Asia)
Junonia orithya (Asia)
Junonia hierta (Africa & Asia)
Junonia oenone (Africa)
Junonia orithya (Africa)
Junonia sophia (Africa)
Junonia westermanni (Africa)
New World Junonia
(J. vestina not included in the analysis)

Taxonomic history of Junonia

In the New World, there are 10 currently recognized species (Brévignon, 2004, Brévignon, 2009, Gemmell et al., 2014) (Table 1-1) and there are 19 *Junonia* species that occur in the Old World (Table 1-2) (Wahlberg et al., 2005). There have been several hypotheses about the origins of *Junonia* in the New World, based on phenotypic characters (Fig. 1-1). Based on similarities of wing color pattern and male genital morphology, Forbes (1928, 1947, Fig. 1-1A) hypothesized that *J. villida* from the Asian-Pacific region was the sister taxon to *J. vestina* which occurs in South America, and which is in turn the sister taxon to all the other New World species. Eliot (1946, 1947, Fig. 1-1B) referred to features of wing color patterns and biogeographical considerations (especially the derivation of the African and South American continents from a fragmenting Gondwanaland supercontinent) to propose *J. orithya* from Africa as the sister taxon to New World *Junonia*. Corbet (1948, Fig. 1-1C) consulted features of wing color patterns and flight behavior and also proposed that *J. orithya* was the likely candidate for sister taxon to the New World *Junonia*, but from the Asia-Pacific region and not Africa as previously suggested (Eliot, 1946, Eliot, 1947).

The first molecular investigations of *Junonia* phylogeny used 3 partial DNA sequences: mitochondrial *cox1* barcodes, and nuclear *wingless*, and nuclear *ef1-alpha*, to find that the New World *Junonia* were monophyletic, but included only 3 New World species and were not able to resolve which taxon might the sister taxon to the New World *Junonia* species (Kodandaramaiah & Wahlberg, 2007, Kodandaramaiah, 2009) (Fig. 1-1D). All 3 of the taxa identified as possible sister taxa based on phenotypic considerations by prior workers (Fig1-1A-C) remained possible sister taxa in the molecular phylogenetic analysis. Gemmell and Marcus (2015) reproduced these results in a phylogenetic analysis of *cox1* barcodes that included all of the New World *Junonia*

taxa, but observed that the bootstrap support for New World monophyly was extremely weak. Gemmell and Marcus (2015) proceeded to conduct a haplotype network analysis of *cox1* haplotypes that suggested that the *cox1* haplotypes in *J. villida* appear to have greater sequence similarity to the New World *Junonia* than the *cox1* haplotypes found in *J. orithya* from either Africa or Asia. Gemmell and Marcus (2015) also observed that the *J. villida* haplotype grouped most closely with New World haplotype group A2 and that both of these haplotypes were very distinct from New World haplotype group B. These observations led Gemmell and Marcus (2015) to suggest that the New World *Junonia* haplotypes may not be monophyletic and may have arisen through multiple invasions of the New World.

Table 1-1. New World *Junonia* species with geographic and altitudinal ranges (Gemmell & Marcus, 2015).

Species	**Species Authority**	**Geographic Range**	**Altitudinal Range**
J. coenia	Hübner, 1822	N. America	0 – 1000 m
J. divaricata	C. & R. Felder, 1867	Guianas, S. America	0 – 1000 m
J. evarete	Cramer, 1779	S. America, S. parts N. America	0 – 1500 m
J. genoveva	Cramer, 1780	S. America, S. parts N. America	0 – 2100 m
J. litoralis	Brevignon, 2009	Atlantic Coast S. America	0 – 50 m
J. neildi	Brevignon, 2004	Caribbean	0 – 50 m
J. nigrosuffusa	Barnes & McDunnough,	Mexico, SW USA	0 – 1000 m
J. vestina	C. & R. Felder, 1867	South American Andes	1900 – 3500 m
J. wahlbergi	Brevignon, 2008	Guianas, S. America	0 – 1000 m
J. zonalis	C. & R. Felder, 1867	Caribbean	0 – 1000 m
Forms of *Junonia* of indeterminate taxonomic status[1]:			
J. sp. flirtea	Fabricius, 1793	Argentina S. America	0 – 1000 m
J. sp. hilaris	C. & R. Felder, 1867	Argentina S. America	0 – 1000 m

[1] *Junonia sp. flirtea* was described as a subspecies of *J. evarete* and *J. sp. hilaris* was described as a subspecies of *J. genoveva*. However, recent work suggests that these forms are genetically more closely related to each other than either form is to the nominate subspecies of *J. evarete* and *J. genoveva* (Borchers & Marcus, 2014, Gemmell et al., 2014). Thus, their taxonomic status is ambiguous.

Table 1-2. Old World *Junonia* species with geographic ranges (Igarishi & Fukada, 1997, Vane-Wright & Tennent, 2011, Mecenero et al., 2013, Savela, 2015). (all low elevation 0 – 1000 m)

Species	Species Authority	Geographic Range
J. almana	Linnaeus, 1758	India, SE Asia, Indonesia, Philippines
J. ansorgei	Rothschild, 1899	Central Africa
J. artaxia	Hewitson, 1864	Central Africa
J. atlites	Linnaeus, 1763	India, SE Asia, Indonesia, Philippines
J. cymodoce	Cramer, 1777	West Africa
J. erigone	Cramer, 1775	Indonesia, New Guinea
J. gregorii	Butler, 1896	East Africa
J. hedonia	Linnaeus, 1764	Indonesia, Philippines, New Guinea
J. hierta	Fabricius, 1798	Africa, Middle East, India, SE Asia
J. iphita	Cramer, 1779	India, SE Asia, Indonesia
J. lemonias	Linnaeus, 1758	India, SE Asia, Philippines
J. natalica	C. & R. Felder,	Sub-Saharan Africa
J. oenone	Linnaeus, 1758	Sub-Saharan Africa, Arabian Peninsula
J. orithya	Linnaeus, 1758	Africa, Middle East, India, SE Asia, Indonesia, Australia
J. sophia	Fabricius, 1793	Central Africa
J. terea	Drury, 1773	Sub-Saharan Africa
J. touhilimasa	Vuillot, 1892	East Africa
J. westermanni	Westwood, 1870	Sub-Saharan Africa
J. villida	Fabricius, 1787	Chagos Islands, Indonesia, New Guinea, Australia, Micronesia, Polynesia

Dispersal in butterflies

Compared with other insect groups, most butterfly species are generally poor at natural long-distance dispersal due to their relatively large body size and relatively poor vagility (Miller, 1984). Almost all butterfly species are confined to a single continent, with the vast majority confined to only a portion of a continent (Sbordoni & Forestiero, 1998). The butterfly faunas of North America, Europe, and Asia have many affinities, since the same butterfly genera are frequently found on all 3 continents, and are often considered to be a single Holarctic fauna (Hoskins, 2015). The butterfly faunas of the southern continents, Africa, Australia, and South America, are distinct from the Holarctic fauna, and from each other. Relatively few genera are shared between South America and either Africa or Australia (Hoskins, 2015).

The butterfly fauna of isolated oceanic islands is generally species-poor when compared with both mainland butterfly faunas and with the island faunas of other insect groups (Miller, 1984). For example, there are only two native butterfly species in Hawaii (Roderick & Gillespie, 1998): *Vanessa tameamea* Eschscholtz 1821 and *Udara blackburni* Toxopeus 1928. By contrast, there are over 300 native species of *Drosophila* (Ringo, 1977), 128 native species in the beetle tribe Platynini (Roderick & Gillespie, 1998), and over 300 native (and endemic) species in the moth genus *Hyposmocoma* (Rubinoff, 2008). When butterflies do reach isolated oceanic islands, they frequently produce endemic species because of very low gene flow between island and mainland source populations (Adler & Dudley, 1994). As a result, isolated oceanic islands often have a small butterfly fauna with a disproportionate number of endemic butterfly species (Adler & Dudley, 1994, Roderick & Gillespie, 1998, Patrick & Patrick, 2012).

Among the butterfly families, Nymphalid butterflies are among the best at long-range dispersal. Among the ten species shared (naturally, without human intervention) between Europe and North America, 7 species are Nymphalid butterflies (Carter, 1982, Brock & Kaufman, 2003). Within the Nymphalidae, there are several genera that appear to have been especially effective at unaided long-range dispersal. Often among the only native butterfly genera present on remote oceanic islands belong to a small number of Nymphalid genera: *Vanessa, Hypolimnas, Danaus,* and *Junonia* (Roderick & Gillespie, 1998, Vane-Wright & Tennent, 2011, Patrick & Patrick, 2012). Even with human facilitation, it is often species from these genera that have been brought to isolated islands (e.g. Hawaii: 4 *Vanessa* (1 native, 3 introduced), 1 introduced *Danaus*) (Denny & Jamieson, 2001). *Hypolimnas misippus* has extended its range to the New World since European contact, possibly in association with shipping between Africa and the Caribbean (Brown & Heineman, 1972). *Danaus plexippus* has extended its range from North America to

Bermuda, the Azores, the Canary Islands, Hawaii, Australia, and New Zealand during recorded history, probably facilitated by human transport (CBIF, 2006).

The dispersal of *Junonia* species is known to have occurred between continents: the genus originated in Africa yet was able to disperse to south Asia, the New World, and from Asia back to Africa (Kodandaramaiah & Wahlberg, 2007). *Junonia villida* in particular has dispersed to islands stretching from the Chagos Islands of the Indian Ocean all the way to the Gambier Islands in French Polynesia in the Pacific Ocean (a linear distance of 15,500 km, 154º longitude, or 43% of the earth's circumference) (Vane-Wright & Tennent, 2011). Three other widespread *Junonia* species are *J. orithya* (Japan, Taiwan, Philippines, New Guinea, Australia, Southeast and South Asia, Arabia, Madasgascar, sub-Saharan Africa) *J. hierta* (Southeast and South Asia, Andaman Islands, Nicobar Islands, Arabia, Madasgascar, sub-Saharan Africa), (Larsen, 1999) and *J. coenia* (much of North America (Yucatan to Southern Canada), Cuba, Bahamas, Bermuda) (Scott, 1986).

Finally, some *Junonia* species are migratory and are apparently good at long-range dispersal. *Junonia villida* is migratory within Australia (Braby, 2000), and appears to have dispersed on several occasions from Australia to New Zealand, where it is apparently unable to complete its life cycle (Harris, 1988). *Junonia coenia* recolonizes almost all of the continental United States every summer from overwintering sites along the Gulf Coast, reaching as far north as Manitoba in some years (Klassen et al., 1989, Shapiro, 1991). Similar seasonal migrations apparently occur in *J. sp. flirtea* and *J. sp. hilaris* in Argentina in the summer (Williams, 1930). These features of *Junonia* dispersal biology may have had an important impact on *Junonia* biogeography and diversification (Vane-Wright & Tennent, 2011). *Junonia* is primarily a tropical genus, so long-range dispersal to the New World from either Africa or the Asian Pacific

region is suspected as having played a role in the colonization of the Western Hemisphere (Fig. 1-1), perhaps more than once (Gemmell & Marcus, 2015).

Junonia butterflies as model organisms

Members of the genus *Junonia* have been used as model organisms in many different fields of biology. *Drosophila melanogaster* is the best-studied insect model organism but its development is in many respects highly derived and not necessarily representative of other insects, and the ecology of *Drosophila melanogaster* Meigen, 1830 in its natural or native environment is almost entirely unknown (Marcus, 2005). *Junonia* has embryonic development that is more typical of most other holometabolous insects (Truman & Riddiford, 1999), and many aspects of its ecology are well studied (Clark, 1932, Scott, 1975, Kodandaramaiah, 2011). Almost all *Junonia* species inhabit tropical or temperate regions, and generally prefer low elevation open habitats with little vertical structure such as prairies, grasslands, sand dunes, and vacant lots (Glassberg et al., 2000).

The evolution of larval host plant preferences has also been a topic of interest in this genus, and the larval host plants of many *Junonia* species have been investigated. Across the genus, *Junonia* larvae feed on many different host plant species, nearly all of which contain iridoid glycosides (Bowers, 1984, Bowers & Stamp, 1997). These compounds are distasteful to vertebrate predators (Bowers & Farley, 1990), and are sequestered in the bodies of *Junonia* larvae during development (Bowers & Puttick, 1989) but are purged at pupation and the adults are palatable to vertebrate predators (Bowers & Collinge, 1992). Some *Junonia* species can be identified on the basis of their larval host plant preferences, given that many species only feed on one or two plant species. Out of the seven *Junonia* species that occur in French Guiana,

Martinique, and Guadeloupe, five species have only one larval host plant, and the remaining two species have two larval host plant species (Brévignon, 2003, Brévignon & Brévignon, 2012). *Junonia* adult flight periods can often be predicted based on the growing seasons of the larval host plant of each species (Brévignon, 2003, Brévignon, 2004, Brévignon & Brévignon, 2012). In North America, *J. coenia* also specializes on plants containing iridoid glycoside compounds, but its larval host range includes plants in seven different families (Knerl & Bowers, 2013). However, specific *J. coenia* populations in North America may choose their preferred host plant based on availability in their local habitat (Camara, 1997). There is also an artificial diet containing iridoid glycosides available for *Junonia*, facilitating year-round laboratory culture and this can be used to successfully raise larvae of most New World *Junonia* species to adulthood ((Paulsen, 1996, Ellis & Bowers, 1998), J. M. Marcus, pers. comm.).

Research has also been conducted into the physiological conditions of wing development in *J. coenia*. Wing disks from *J. coenia* can be cultured *in vitro*, increasing the amount of manipulation that can be performed in the laboratory (Miner et al., 2000, Nijhout & Grunert, 2002). By manipulating nutrients and growth hormones *in vitro*, it was found that wing disks in *J. coenia* are committed at the pupal stage one cell at a time, with cell commitment continuing throughout the entire wing disk, as well as the identities of many of the hormones involved in this process (Kremen & Nijhout, 1989, Kremen & Nijhout, 1998, Nijhout & Grunert, 2002). *Junonia* wing discs can be genetically modified to transiently express transgenes by somatic transformation using baculovirus, sinbis, or vaccinia virus (Lewis et al., 1999, Lewis & Brunetti, 2006, Dhungel et al., 2013). Finally, *Junonia* species can be genetically modified by germ line transformation, which has been used to create lines of butterflies that express enhanced yellow

fluorescent protein (Beaudette et al., 2014) and could also be used to manipulate expression of endogenous gene products.

Due to their prominent eyespots and bright colors, many studies have investigated the development and physiology of eyespots and color patterns in *Junonia*. Cells at the center of each eyespot focus are responsible for patterning the surrounding eyespots, and even the physical movement of foci cells into other areas of the wing still creates at least partial eyespot expression in the new wing tissue (Nijhout, 1980). Studies of gene expression in *J. coenia* wings have revealed a large number of gene products that are expressed in spatial and temporal patterns that are consistent with roles in eyespot development (Carroll et al., 1994, Keys et al., 1999, Brunetti et al., 2001, Reed & Serfas, 2004, Martin & Reed, 2010). *Junonia* has been used as an investigative tool for eyespot evolution on butterfly wings, and they are a good model given that eyespots are present in all *Junonia* species, but show a variety of different phenotypes (Kodandaramaiah, 2009). *Junonia* species may have had lost and regained eyespots repeatedly throughout its evolutionary history (Kodandaramaiah, 2009).

Molecular phylogenetic approach

Despite all of the mechanistic research that has been conducted in *Junonia*, most studies have been conducted on *J. coenia* and in a few other New World species, making the phylogeny of this part of the genus of particular interest. The early molecular phylogenetic treatments of the genus included only 3 New World species (Kodandaramaiah & Wahlberg, 2007, Kodandaramaiah, 2009) (Fig. 1-1D). More recent phylogenetic work had improved species sampling, but has been largely restricted to analysis of *cytochrome c oxidase I (cox1)* barcodes (Pfeiler et al., 2012), which are not effective in distinguishing *Junonia* species (Brévignon &

Brévignon, 2012, Gemmell et al., 2014, Gemmell & Marcus, 2015). In order to trace the evolution of all of these biological processes and their effects on phenotype within the New World *Junonia,* it is necessary to produce a robust phylogenetic framework for this complex part of the genus.

One promising new technique for molecular phylogenetics is to develop data sets from complete mitochondrial genomes (Hao et al., 2013, Gillett et al., 2014, Timmermans et al., 2014, Wu et al., 2014). These data sets are readily generated from next generation sequencing of DNA extracts. In insects, mitochondrial genomes consist of more than 15,000 bp, providing a rich source of informative characters for phylogenetic analysis. Prior to my thesis project, there were only two *Junonia* mitochondrial genomes available, both from the Old World (Shi et al., 2013, Wu et al., 2014), I have expanded upon this baseline by sequencing 14 complete mitochondrial genomes, ten from the New World and four additional mitochondrial genomes from the Old World. I also partially sequenced two further New World mitochondrial genomes, with greater than 90% of the genome sequenced. Specimens to be sequenced were selected based on taxonomy, geography, quality of DNA preservation, and prior information about the specimen obtained from *cox1* barcodes (Gemmell & Marcus, 2015). Determining the structure and level of conservation of synteny among *Junonia* mitochondrial genomes will help define the phylogenetic approaches that will employ these data.

Of the mitochondrial genomes that I sequenced, one of particular note was from *J. vestina*, which only occurs from 1900-3500 m elevation in the Andes Mountains ((Forbes, 1928, Neild, 2008, Pfeiler et al., 2012) K. R. Willmott, pers. comm.). Most other New World *Junonia* are rarely found above 2000 m elevation (Neild, 2008), K. R. Willmott, pers. comm.). *Junonia vestina* is also unusual among New World *Junonia* species because it has a distinct *cox1*

haplotype (A1) (Pfeiler et al., 2012). The remaining New World species share and are polymorphic for two alternative haplotype groups (A2 and B) both of which occur in most species (Gemmell & Marcus, 2015). The distinctiveness of the *J. vestina* haplotype may be connected to its high elevation environment, which is characterized by low oxygen concentrations and low temperatures (Hassanin et al., 2009). The mitochondrial genomes of many high elevation vertebrates (especially mammals) have been studied and many different mitochondrial genes have been identified as possibly being under selection at high elevation (Luo et al., 2013). However, because prior studies have focused mainly on mammals at high elevations, it has been difficult to differentiate between the effects of high altitude conditions on selection for mitochondrial function, versus the effects of endothermy at high elevation on mitochondrial function (Hassanin et al., 2009). By comparing *J. vestina* to closely related low elevation *Junonia* mitochondrial genomes, it may be possible to identify portions of the mitochondrial genome that have been selected for performance at high elevation conditions. By placing the New World *Junonia* in a phylogenetic context, it becomes a more powerful model system that will be informative not only for understanding adaptive evolution in this group of butterflies, but also for understanding high elevation adaptation of mitochondria more generally, and to separate the effects of high elevation from the effects of endothermy in high elevation mammals.

Objectives and rationale of the thesis

1. To characterize the detailed structure of a complete *Junonia* mitochondrial genome and to compare it with previously sequenced butterfly mitochondrial genomes in order predict

the expected degree of conservation of synteny among the mitochondrial genomes within this genus. (Chapter 2)

2. To sequence fifteen Old and New World *Junonia* mitochondrial genomes and use them in concert with previously sequenced *Junonia* mitochondrial genomes to create a robust phylogeny of the mitochondrial haplotypes found in the New World members of genus *Junonia* in order to determine how many invasions of the New World occurred and from where they originated. (Chapter 3)
3. To use the amount of divergence between whole mitochondrial genome sequences to establish time frames for the origin of each the haplotype groups that occur in the New World. (Chapter 3)
4. To investigate the differences between *J. vestina* and other members of the genus in order to determine potential areas of sequence divergence relate to high altitude adaptation. (Chapter 4)
5. To determine if mitochondrial mutations related to high altitude adaptation in *J. vestina* also occur in other high elevation species, and whether there are any diseases associated at the same amino acid positions within the same genes in human populations. Pathology associated with specific amino acid positions in mitochondrial gene products can help associate biological function with specific amino acids in a peptide sequence. This can give clues to the possible functional consequences of mutations at the same position in high elevation butterflies. (Chapter 4)

Literature Cited

Adler, G.H. and Dudley, R., 1994. Butterfly biogeography and endemism on tropical Pacific Islands. Biol. J. Linn. Soc. 52, 151-162.

Beaudette, K., Hughes, T.M. and Marcus, J.M., 2014. Improved injection needles facilitate germ-line transformation of the buckeye butterfly *Junonia coenia*. Biotechniques 56, 142-144. doi: 10.2144/000114147.

Borchers, T.E. and Marcus, J.M., 2014. Genetic population structure of buckeye butterflies (*Junonia*) from Argentina. Syst. Ent. 39, 242-255. doi: 10.1111/syen.12053.

Bowers, M.D., 1984. Iridoid Glycosides and Host-Plant Specificity in Larvae of the Buckeye Butterfly, *Junonia coenia* (Nymphalidae). J. Chem. Ecol. 10, 1567-1577.

Bowers, M.D. and Collinge, S.K., 1992. Fate of iridoid glycosides in different life stages of the buckeye, *Junonia coenia* (Lepidoptera: Nymphalidae). J. Chem. Ecol. 18, 817-831.

Bowers, M.D. and Farley, S., 1990. The behavior of gray jays, *Perisoreus canadensis*, towards palatable and unpalatable Lepidoptera. Anim. Behav. 39, 699-705.

Bowers, M.D. and Puttick, G.M., 1989. Iridoid glycosides and insect feeding preferences - gypsy moths (*Lymantria dispar*, Lymantriidae) and buckeyes (*Junonia coenia*, Nymphalidae). Ecol. Entomol. 14, 247-256.

Bowers, M.D. and Stamp, N.E., 1997. Effect of hostplant genotype and predators on iridoid glycoside content of pupae of a specialist insect herbivore, *Junonia coenia* (Nymphalidae). Biochemical Systematics and Ecology 25, 571-580.

Braby, M.F., 2000. Butterflies of Australia: Their identification, biology and distribution, CSIRO Publishing, Collingwood, Australia.

Brévignon, C., 2003. Les papillon diurnes de la Guadeloupe. Lambillionea CIII, 1-29.

Brévignon, C., 2004. Description de deux nouvelle sous-espèces Guadeloupeennes du genere *Junonia* Hübner, 1819 (Lepidoptera, Nymphalidae, Nymphalinae). Lambillionea CIV, 72-80.

Brévignon, L. and Brévignon, C., 2012. Nouvelles observations sur le genre *Junonia* en Guyane Française. (Lepidoptera: Nymphalidae) (Seconde partie). Lépidoptères de Guyane 7, 8-35.

Brock, J.P. and Kaufman, K., 2003. Butterflies of North America, Houghton Mifflin New York.

Brown, F.M. and Heineman, B., 1972. Jamaica and its Butterflies., E. W. Classey, Ltd., London.

Brunetti, C.R., Selegue, J.E., Monteiro, A., French, V., Brakefield, P.M. and Carroll, S.B., 2001. The generation and diversification of butterfly eyespot colour patterns. Curr. Biol. 11, 1578-1585.

Camara, M.D., 1997. A recent host range expansion in *Junonia coenia* Huebner (Nymphalidae): Oviposition preference, survival, growth, and chemical defense. Evolution 51, 873-884.

Carroll, S.B., Gates, J., Keys, D.N., Paddock, S.W., Panganiban, G.E.F., Selegue, J.E. and Williams, J.A., 1994. Pattern formation and eyespot determination in butterfly wings. Science 265, 109-114.

Carter, D., 1982. Butterflies and Moths in Britain and Europe, Pan Books, London.

CBIF, 2006. Butterflies of Canada. http://www.cbif.gc.ca/spp_pages/butterflies/index_e.php.

Clark, A.H., 1932. Butterflies of the District of Columbia and vicinity. Bull. U.S. Nat. Mus. 157, 1-338.

Corbet, A.S., 1948. Papers on Malaysian Rhopalocera. V. The conspecificity of the American *Precis lavinia* (Cramer) with the Oriental *Precis orithya* (Cramer). Entomologist 81, 54-56.

Denny, J. and Jamieson, D.G., 2001. Hawaii's Butterflies and Moths: An Identification Guide to Easily Observed Species, Mutual, Honolulu, Hawaii.

Dhungel, B., Ohno, Y., Matayoshi, R. and Otaki, J.M., 2013. Baculovirus-mediated gene transfer in butterfly wings *in vivo*: an efficient expression system with an anti-gp64 antibody. BMC Biotechnol. 13, 27. doi: 10.1186/1472-6750-13-27.

Eliot, N., 1946. Continental drift and *Precis lavinia.* Entomologist 79, 225-228.

Eliot, N., 1947. More on continental drift, *Precis lavinia* Hb. and *P. villida* F. Entomologist 80, 230-234.

Ellis, A. and Bowers, M.D., 1998. Effects of hostplant species and artificial diet on growth of buckeye (*Junonia coenia*) and painted lady (*Vanessa cardui*) caterpillars (Nymphalidae). J. Lepid. Soc. 52, 73-83.

Forbes, W.T.M., 1928. Variation in *Junonia lavinia* (Lepidoptera, Nymphalidae). J. N.Y. Ent. Soc. 36, 306-321.

Forbes, W.T.M., 1947. Buckeyes and Wegener. Entomologist 80, 56-58.

Gemmell, A.P., Borchers, T.E. and Marcus, J.M., 2014. Genetic Population Structure of Buckeye Butterflies (*Junonia*) from French Guiana, Martinique, and Guadeloupe. Psyche 2014, 1-21. doi: 10.1155/2014/897596.

Gemmell, A.P. and Marcus, J.M., 2015. A tale of two haplotype groups: The origin and distribution of divergent New World *Junonia COI* haplotypes. Syst. Ent. 40, 532-546. doi: 10.1111/syen.12120.

Gillett, C.P.D.T., Crampton-Platt, A., Timmermans, M.J.T.N., Jordal, B.H., Emerson, B.C. and Vogler, A.P., 2014. Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). Mol. Biol. Evol. 31, 2223-2237.

Glassberg, J., Minno, M.C. and Calhoun, J.V., 2000. Butterflies through binoculars: A field, finding, and gardening guide to butterflies in Florida, Oxford, New York.

Hao, J., Sun, M., Shi, Q., Sun, X., Shao, L. and Yang, Q., 2013. Complete Mitogenomes of *Euploea mulciber* (Nymphalidae: Danainae) and *Libythea celtis* (Nymphalidae: Libytheinae) and Their Phylogenetic Implications. ISRN Genomics 2013, 14.

Harris, A.C., 1988. A large migration of the Australian meadow argus butterfly *Junonia villida calybe* (Lepidopera: Nymphalidae) to southern New Zealand. New Zealand Entomologist 11, 67-68.

Hassanin, A., Ropiquet, A., Coulous, A. and Cruaud, C., 2009. Evolution of the mitochondrial genome in mammals living at high altitude: New insights from a study of the tribe Caprini (Bovidae, Antilopinae). J. Mol. Evol. 68, 293-310 DOI 10.1007/s00239-009-9208-7.

Hoskins, A., 2015. Butterflies of the World, New Holland, Sydney, Australia.

Igarishi, S. and Fukada, H., 1997. The Life Histories of Asian Butterflies, Volume 1, Tokai University Press.

Keys, D.N., Lewis, D.L., Selegue, J.E., Pearson, B.J., Goodrich, L.V., Johnson, R.J., Gates, J., Scott, M.P. and Carroll, S.B., 1999. Recruitment of a *hedgehog* regulatory circuit in butterfly eyespot evolution. Science 283, 532-534.

Klassen, P., Westwood, A.R., Preston, W.B. and McKillop, W.B., 1989. Butterflies of Manitoba, University of Manitoba Press, Winnipeg, Canada.
Knerl, A. and Bowers, M.D., 2013. Incorporation of an introduced weed into the diet of a native butterfly: Consequences for preference, performance and chemical defense. J. Chem. Ecol. 39, 1313-1321. doi: 10.1007/s10886-013-0355-3.
Kodandaramaiah, U., 2009. Eyespot evolution: Phylogenetic insights from *Junonia* and related butterfly genera (Nymphalidae: Junoniini). Evol. Dev. 11, 489-497.
Kodandaramaiah, U., 2011. The evolutionary significance of butterfly eyespots. Behav. Ecol. 22, 1264-1271.
Kodandaramaiah, U. and Wahlberg, N., 2007. Out-of-Africa origin and dispersal-mediated diversification of the butterfly genus *Junonia* (Nymphalidae: Nymphalinae). J. Evol. Biol. 20, 2181-2191.
Kremen, C. and Nijhout, H.F., 1989. Juvenile hormone controls the onset of pupal commitment in the imaginal disks and epidermis of *Precis coenia* (Lepidoptera: Nymphalidae). J. Insect Physiol. 35, 603-612.
Kremen, C. and Nijhout, H.F., 1998. Control of pupal commitment in the imaginal disks of *Precis coenia* (Lepidoptera: Nymphalidae). J. Insect Physiol. 44, 287-296.
Larsen, T., 1999. Butterflies of West Africa - origins, natural history, diversity, and conservation. Draft systematic part ed. Larsen.
Lewis, D.L. and Brunetti, C.R., 2006. Ectopic transgene expression in butterfly imaginal wing discs using vaccinia virus. Biotechniques 40, 48-54.
Lewis, D.L., DeCamillis, M.A., Brunetti, C.R., Halder, G., Kassner, V.A., Selegue, J.E., Higgs, S. and Carroll, S.B., 1999. Ectopic gene expression and homeotic transformations in arthropods using recombinant Sindbis viruses. Curr. Biol. 9, 1279-1287.
Luo, Y., Yang, X. and Gao, Y., 2013. Mitochondrial DNA response to high altitude: a new perspective on high-altitude adaptation. Mitochondrial DNA 24, 313-319. doi: 10.3109/19401736.2012.760558.
Marcus, J.M., 2005. Jumping genes and AFLP maps: Transforming Lepidopteran color pattern genetics. Evol. Dev. 7, 108-114.
Martin, A. and Reed, R.D., 2010. *wingless* and *aristaless2* define a developmental ground plan for moth and butterfly wing pattern evolution. Mol. Biol. Evol. 27, 2864-2878 doi: 10.1093/molbev/msq173.
Mecenero, S., Ball, J.B., Edge, D.A., Hamer, M.L., Henning, G.A., Kruger, M., Pringle, E.L., Terblanche, R.F. and Williams, M.C., 2013. Conservation Assessment of Butterflies of South Africa, Lesotho and Swaziland – Red List and Atlas, Saftronics and the Animal Demography Unit, University of Cape Town.
Miller, S.E., 1984. Butterflies of the California Channel Islands. J. Res. Lep. 23, 282-296.
Miner, A.L., Rosenberg, A.J. and Nijhout, H.F., 2000. Control of growth and differentiation of the wing imaginal disk of *Precis coenia* (Lepidoptera: Nymphalidae). J. Insect Physiol. 46, 251-258.
Neild, A.F.E., 2008. The Butterflies of Venezuela, Part 2: Nymphalidae II (Acraeinae, Libytheinae, Nymphalinae, Ithomiinae, Morphinae), Meridian, London.
Nijhout, H.F., 1980. Pattern formation on lepidopteran wings: Determination of an eyespot. Dev. Biol. 80, 267-274.
Nijhout, H.F. and Grunert, L.W., 2002. Bombyxin is a growth factor for wing imaginal disks in Lepidoptera. Proc. Nat. Acad. Sci. USA 99, 15446-15450.

Patrick, B. and Patrick, H., 2012. Butterflies of the South Pacific, Otago University Press, Otago, New Zealand.
Paulsen, S.M., 1996. Quantitative genetics of the wing color pattern in the buckeye butterfly (*Precis coenia* and *Precis evarete*): evidence against the constancy of G. Evolution 50, 2360-2372.
Pfeiler, E., Johnson, S. and Markow, T.A., 2012. DNA barcodes and insights into the relationships and systematics of buckeye butterflies (Nymphalidae: Nymphalinae: *Junonia*) from the Americas. J. Lepid. Soc. 66, 185-198.
Reed, R.D. and Serfas, M.S., 2004. Butterfly wing pattern evolution is associated with changes in a Notch/Distal-less temporal pattern formation process. Curr. Biol. 14, 1159-1166.
Ringo, J.M., 1977. Why 300 species of Hawaiian *Drosophila*? The sexual selection hypothesis. Evolution 31, 694-696.
Roderick, G.K. and Gillespie, R.G., 1998. Speciation and phylogeography of Hawaiian terrestrial arthropods. Mol. Ecol. 7, 519-531.
Rubinoff, D., 2008. Phylogeography and ecology of an endemic radiation of Hawaiian aquatic case-bearing moths (Hyposmocoma: Cosmopterigidae). Philos. Trans. R. Soc. Lond., B, Biol. Sci. 363, 3459-3465.
Savela, M., 2015. Lepidoptera and some other life forms. http://www.nic.funet.fi/pub/sci/bio/life/intro.html.
Sbordoni, V. and Forestiero, S., 1998. Butterflies of the World, Firefly Books, Richmind Hill, Ontario, Canada.
Scott, J.A., 1975. Movement of *Precis coenia*, a pseudoterritorial submigrant (Lepidoptera-Nymphalidae). J. Anim. Ecol. 44, 843-850.
Scott, J.A., 1986. The butterflies of North America : a natural history and field guide, Stanford University Press, Stanford, CA.
Shapiro, A.M., 1991. Extirpation and recolonization of the buckeye, *Junonia coenia* (Nymphalidae) following the northern California freeze of December, 1990. J. Res. Lep. 30, 209-220.
Shi, Q., Huang, D., Wang, Y. and Hao, J., 2013. The complete mitochondrial genome of Blue Pansy, *Junonia orithya* (Lepidoptera: Nymphalidae: Nymphalinae). Mitochondrial DNA 0, 1-2. doi: 10.3109/19401736.19402013.19823182.
Timmermans, M.J.T.N., Lees, D.C. and Simonsen, T.J., 2014. Towards a mitogenomic phylogeny of Lepidoptera. Mol. Phylogenet. Evol. 79, 169-178.
Truman, J.W. and Riddiford, L.M., 1999. The origins of insect metamorphosis. Nature 401, 447-452.
Vane-Wright, R.I. and Tennent, W.J., 2011. Colour and size variation in *Junonia villida* (Lepidoptera, Nymphalidae): subspecies or phenotypic plasticity? Syst. and Biodivers. 9, 289-305.
Wahlberg, N., Brower, A.V.Z. and Nylin, S., 2005. Phylogenetic relationships and historical biogeography of tribes and genera in the subfamily Nymphalinae (Lepidoptera: Nymphalidae). Biol. J. Linn. Soc. 86, 227-251.
Williams, C.B., 1930. The Migration of Butterflies, Oliver and Boyd, Edinburgh.
Wu, L.-W., Lin, L.-H., Lees, D. and Hsu, Y.-F., 2014. Mitogenomic sequences effectively recover relationships within brush-footed butterflies (Lepidoptera: Nymphalidae). BMC Genomics 15, 468 doi:10.1186/1471-2164-15-468.

Chapter 2: The complete mitochondrial genome of the Lemon Pansy, Junonia lemonias (Lepidoptera: Nymphalidae: Nymphalinae)[1]

[1] This chapter has been previously published as:
McCullagh, B. S. and J. M. Marcus. 2015. The complete mitochondrional genome of Lemon Pansy, *Junonia lemonias* (Lepidoptera: Nymphalidae: Nymphalinae). Journal of Asia-Pacific Entomology 18: 749-755. doi: 10.1016/j.aspen.2015.08.006 BSM conceived the research, collected the data, and wrote the manuscript.

Abstract

Junonia lemonias (Lepidoptera: Nymphalidae: Nymphalinae), the lemon pansy, is a nymphalid butterfly found throughout South and Southeast Asia. Its wings display seasonal polyphenism in the size and distinctiveness of their wing color patterns, varying considerably between wet and dry seasons. In both seasonal forms, two prominent eyespots are found on both the forewings and hindwings. The common name for this butterfly comes from the bright lemon-yellow coloring found around the eyespots and wing margins. The complete mitochondrial genome of *J. lemonias* is 15,230 bp long. There are 37 genes, 13 of which are protein-coding. The mitogenome shares many features and is similar in length to other published *Junonia* mitochondrial genomes. There are 22 tRNAs present in the mitochondrial genome, along with two rRNAs and a control region consisting of 333 bp. The mitochondrial genome is AT-rich with 80.40% AT nucleotides. The majority coding sequence strand has a nucleotide skew towards T and C bases. Fifteen intergenic spacers are present, along with 11 regions of sequence overlap. Many conserved features of the mitogenome found across Nymphalidae are also present in *J. lemonias.* Widespread mitochondrial genome synteny occurs in this family, allowing entire mitogenomes to be aligned for phylogenetic analysis. Parsimony and maximum likelihood phylogenetic analysis of both the entire mitochondrial genome and first and second codons of mitochondrial protein-coding genes suggest that *Junonia* and the Nymphalinae are monophyletic and that the Nymphalinae may be the sister group to a clade containing the nymphalid subfamilies Apaturinae, Biblidinae, and Cyrestinae.

Keywords: Nymphalidae, Nymphalinae, *Junonia lemonias*, mitochondrial genome, Next generation sequencing

Introduction

The butterflies of the genus *Junonia* are an important model system for the study of wing color pattern evolution and development (Carroll et al., 1994, Kodandaramaiah, 2009, Otaki, 2012, Martin & Reed, 2014), insect physiology (Nijhout, 2010, Dhungel & Otaki, 2013), and larval host plant evolution (Knerl & Bowers, 2013). The genus has also been the focus of phylogenetic (Kodandaramaiah & Wahlberg, 2007) and biogeographic studies (Vane-Wright & Tennent, 2011, Gemmell & Marcus, 2015). An impressive array of experimental tools have been developed to study the genetics and development of *Junonia* butterflies (Nijhout & Grunert, 2002, Dhungel et al., 2013, Beaudette et al., 2014), but genomic characterization of all the species in this genus is still in early stages (Shi et al., 2013, Wu et al., 2014).

Junonia lemonias (the lemon pansy) Linnaeus, 1758 is a common Asian species that ranges from the Indian subcontinent and Ceylon to south-eastern China, Indo-China, Taiwan and the Philippines (Gu & Chen, 1997, Igarishi & Fukada, 1997, Kunte, 2000). Overall, the lemon pansy imago has brown ground coloration with two eyespots on both the dorsal and ventral surfaces of the forewings and hindwings (Gu & Chen, 1997, Kodandaramaiah, 2009). The prominent lemon-yellow rings around each of the eyespots and yellow markings near the wing margins are the source of the scientific and common name for the species (Linnaeus, 1758, Kunte, 2000). The preferred habitats of *J. lemonias* include grasslands, gardens, open woodlands, and scrub (Kunte, 2000, Rane & Ranade, 2004). Male *J. lemonias* patrol mating territories, often centered on a patch of bare earth, and chase trespassing butterflies that attempt to cross a territory (Kunte, 2000), a pattern of behavior that is similar to that observed in many other Old and New World *Junonia* species (Brown & Heineman, 1972, Turner & Parnell, 1985, Rutowski, 1991). In India, adult *J. lemonias* have been observed to fly from June to November (Rane & Ranade, 2004) and exhibit seasonal polyphenism on their wing surfaces with large, distinct, vivid markings during the wet season and smaller, more subdued indistinct markings during the dry season (Kunte, 2000). This pattern of seasonal polyphenism has also been observed in several other species of *Junonia* in Africa and the New World and in the closely related African genus *Precis* (McLeod, 1984, Smith, 1991), but it is absent or reduced in some other *Junonia* species (Emmel & Emmel, 1973).

Junonia are thought to be specialists on plants that contain iridoid glycosides (Richards et al., 2012), and these compounds are known oviposition (Pereyra & Bowers, 1988) and feeding

stimulants (Bowers, 1984) in some *Junonia* species. Larval hosts of *J. lemonias* include native plant species in the genera *Barleria, Goldfussia, Hygrophila, Nelsonia, Phyla,* and *Sida* (Chou, 1994, Igarishi & Fukada, 1997, Kunte, 2000, Ekpo & Etim, 2009), all of which have been reported to contain iridoid glycoside secondary compounds (Jensen & Neilsen, 1985, Jensen et al., 1988, Yu et al., 2007, Jaiswal et al., 2014, Kim et al., 2015). Three additional plant genera, *Lepidagathis, Corchorus*, and *Cannabis,* are also known larval hosts for *J. lemonias,* but are not yet known to contain iridoid glycosides (Igarishi & Fukada, 1997, Kunte, 2000). *Lepidagathis* is closely related to *Barleria* and likely produces similar secondary compounds including iridoid glycosides (McDade et al., 2008). The secondary compounds produced by *Corchorus* and *Cannabis* have been studied extensively and do not appear to include iridoid glycosides. However, *Corchorus* produces cycloartane triterpene glucosides (Ahmad et al.) and *Cannabis* produces both terpenoids and cannabinoids, all of which have chemical similarities to iridoid glycosides (Turner et al., 1980, ElSohly & Slade, 2005) and may play a similar role in *Junonia*-larval host interactions. Specialization on host plants with different combinations of iridoid glycosides has been implicated as playing a role in both host race formation (Camara, 1997) and cladogenesis (Gemmell et al., 2014) in this genus.

Mitochondrial DNA barcoding has been used extensively to identify *Junonia* species and to infer phylogeny (Brévignon & Brévignon, 2012, Pfeiler et al., 2012). However, these short (658 bp) DNA fragments have been producing puzzling results. In many cases, these DNA fragments show deep mitochondrial divergences within species (Kodandaramaiah, 2009) and identical or nearly identical shared haplotypes between species (Brévignon & Brévignon, 2012, Pfeiler et al., 2012, Borchers & Marcus, 2014). This is possibly due to the lack of sequence evolution as a result of recent divergence, failure of coalescence, or lateral transfer events and organelle capture (Borchers & Marcus, 2014, Gemmell et al., 2014). By assembling additional mitochondrial genome sequences from multiple *Junonia* species, conducting further phylogenetic analyses, and comparing the resulting phylogenies with trees derived from other types of data, it should be possible to distinguish between these alternatives. Next generation sequencing techniques greatly facilitate the sequencing and assembly of complete mitochondrial genomes (Cameron, 2014, Gillett et al., 2014). Here we report the complete annotated mitochondrial genome sequence of *Junonia lemonias* assembled using next generation sequencing techniques to facilitate future studies of phylogenetic relationships and genome

evolution in this genus.

Methods

Sample collection and DNA extraction

A specimen of *J. lemonias* (isolate CNL3) was collected in Yuanjing, Yunnan Province, China on 20 May 2014 by handheld butterfly net. The specimen was killed by pinching, cut in half longitudinally, and immediately stored in RNAlater (Qiagen, Düsseldorf, Germany) at room temperature for two weeks. When the specimen arrived in the laboratory, it was stored at -20ºC. The voucher specimen associated with this mitochondrial genome is labeled with the isolate identifier CNL3 and is currently stored in the Marcus laboratory at the University of Manitoba. It will be donated to an appropriate entomology museum collection when projects that require its physical presence in the laboratory for examination of morphology are complete. DNA was prepared from one leg of this specimen using the DNEasy Blood and Tissue kit (Qiagen) with modifications to the standard animal tissue extraction protocol for total DNA as described previously (Gemmell & Marcus, 2015). Tissue was ground using a ceramic mortar and pestle in 180 µL of tissue lysis buffer ATL. Then, 20 µL of proteinase K (Qiagen, 600 mU/mL) was added to the homogenized sample. The mixture was then incubated in a 55°C water bath for 1 h until the tissue was lysed. The sample was loaded into a Qiagen QiaCube (Qiagen, Düsseldorf, Germany) extraction robot and processed using the standard instrument protocol for purification of total DNA from animal tissue (Gemmell et al., 2014). DNA yield and quality was checked on a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, Delaware, USA) and a Qubit 2.0 fluorometer (Life Technologies, Carlsbad, California, USA). Extracted DNA was stored at -20°C.

DNA preparation and library construction

The extracted *J. lemonias* DNA sample then underwent additional processing at the Next Generation Sequencing Platform facility at the Manitoba Institute of Child Health. The DNA sample was sheared by sonication with an S220 Focused-Ultrasonicator (Covaris, Woburn, Massachusetts, USA) to an average fragment size of 845 bp. Fragment sizes were evaluated using a High Sensitivity DNA chip for the Bioanalyzer 2100 electrophoresis system (Agilent, Santa Clara, California, USA) using the standard manufacturer protocol. A KAPA LTP Library Preparation kit (Kapa Biosystems, Boston, Massachusetts, USA) was used to prepare an

indexed library from the sheared samples for loading onto a MiSeq NextGen Sequencing Instrument (Illumina, San Diego, California, USA). The experiment described here is 1 of 10 indexed libraries that were processed simultaneously on this instrument and represents 10% of the output from 1 run of the instrument. The other nine indexed libraries will be described in Chapter 3. The MiSeq reagent V3 300X2 paired end reagent kit (Illumina) was used to perform the sequencing.

Genome assembly and bioinformatics analysis

Sequence data for *J. lemonias* was aligned to a previously published and annotated *Junonia orithya* complete mitochondrial genome reference sequence (Shi et al., 2013) using Geneious 8.1.3 software (Kearse et al., 2012). After three assembly iterations were performed, no additional fragments were added to the *J. lemonias* contig, so the software automatically terminated the assembly. Once assembly was complete, the annotations from the *J. orithya* reference sequence were transferred with minor adjustments; (Cameron & Whiting, 2008, Cameron, 2014) onto the *J. lemonias* genome sequence using Geneious. The AT and GC content was determined using EndMemo (EndMemo, 2015) and AT/GC skew (Lobry, 1996) were then calculated for the *J. lemonias* mitogenome.

A total of 27 additional mitochondrial genomes from nymphalid butterfly species were identified for comparison purposes on the basis of available full-length mitochondrial genomes in GenBank (Table 2-1). Eight mitogenomes from six different species and five genera within the subfamily Nymphalinae were included. In addition, representatives of eleven other nymphalid subfamilies were included: Apaturinae (8 species), Biblidinae (2 species), and 1 species each from Calinaginae, Charaxinae, Cyrestinae, Danainae, Heliconiinae, Libytheinae, Limenitidinae, Pseudergolinae, and Satyrinae. Specific features to be analyzed (coding sequences, tRNAs, rRNAs, and control region) were retrieved from GenBank sequences using FeatureExtract version 1.2 (Wernersson, 2005). tRNA sequences were analyzed for structure with RNAstructure (Reuters 2010) using default parameters. Sequences from individual protein-coding genes or the entire mitochondrial genome were aligned in Clustal W2 using default settings (Larkin et al., 2007) and the aligned sequences were then converted to NEXUS file format. The aligned sequences for each protein-coding gene were then concatenated into a single NEXUS file to facilitate phylogenetic comparisons with the data set made up of complete

mitochondrial genomes. The best models for maximum likelihood sequence analysis for both the complete mitochondrial genomes and for the data set composed of first and second codon positions of protein coding genes were determined using jModeltest 2.1.7 (Darriba et al., 2012) and a likelihood ratio test (Huelsenbeck & Rannala, 1997): both were found to be General Time Reversible (GTR) models with models for the proportion of invariable sites (I) and for the gamma distribution (G).

Table 2-1. Mitochondrial genome comparison of Nymphalidae species.

Taxon	Mitogenome				Protein-coding genes			tRNA		rRNA		AT-rich region		GenBank Accession no.	References
	Size (bp)	AT (%)	AT-skew	GC skew	CDS length	No. codons	AT (%)	Size (bp)	AT (%)	Size (bp)	AT (%)	Size (bp)	AT (%)		
Nymphalidae															
Nymphalinae															
Junonia lemonias	15230	80.4	-0.008	-0.224	11196	3732	79.02	1474	81.00	2106	84.33	333	94.29	KP941756	This study
Junonia almana	15256	80.26	-0.010	-0.227	11171	3724	78.84	1454	80.54	2119	84.52	185	91.89	NC_024407.1	Wu et al, 2014
Junonia orithya	15214	80.4	-0.009	-0.180	11187	3729	79.17	1461	80.90	2110	83.36	331	94.86	KF199862.1	Shi et al., 2013
Hypolimnas bolina	15260	79.67	-0.018	-0.193	11186	3729	78.11	1460	80.55	2109	83.93	356	93.26	NC_026072.1	Shi et al., 2015
Kallima inachus	15183	80.3	-0.011	-0.228	11197	3733	79.24	1450	80.07	2109	83.64	376	92.02	NC_016196.1	Qin et al., 2012
Kallima inachus	15150	80.6	-0.008	-0.224	11180	3727	79.44	1462	80.51	2118	84.42	331	94.26	HM243591.1	Xu et al., unpublished
Melitaea cinxia	15162	79.96	-0.019	-0.197	11185	3729	78.59	1465	80.61	2105	84.61	332	93.67	HM243592.1	Xu et al., unpublished
Melitaea cinxia	15170	80.01	-0.017	-0.200	11180	3727	78.64	1463	80.72	2108	84.58	338	92.90	NC_018029.1	Xu et al., 2009
Yoma sabina	15330	81.37	-0.009	-0.180	11172	3724	79.99	1456	81.11	2113	85.00	126	92.86	KF590535.1	Wu et al., 2014
Cyrestinae															
Cyrestis thyodamas	15254	81.01	-0.016	-0.188	11190	3730	79.76	1460	80.96	2108	84.82	380	91.58	NC_026071.1	Hao et al., unpublished
Apaturinae															
Apatura ilia	15242	80.45	-0.011	-0.207	11155	3719	78.90	1447	81.48	2109	84.97	403	92.55	NC_016062.1	Chen et al., 2012
Apatura metis	15236	80.44	-0.011	-0.211	11143	3715	78.96	1454	81.36	2112	84.61	394	92.89	NC_015537.1	Zhang et al., 2012
Chitoria ulupi	15279	79.81	-0.026	-0.228	11188	3730	78.22	1458	80.86	2110	84.60	391	90.28	NC_026569.1	Wang et al.,
Sasakia charonda	15244	79.87	-0.007	-0.217	11120	3707	78.22	1458	81.14	2098	84.60	380	91.84	NC_014224.1	Hakozaki et al., unpublished

Sasakia charonda	15233	79.90	-0.007	-0.218	11200	3734	78.37	1449	81.23	2106	84.33	381	91.60	JX119051.1	Wang et al., 2012a
coreana															
Sasakia charonda	15222	79.89	-0.007	-0.218	10795	3599	78.30	1459	81.22	2086	84.51	380	91.84	NC_014223.1	Hakozaki et al., unpublished
kuriyamaensis															
Sasakia funebris	15233	81.23	-0.017	-0.193	11212	3738	79.98	1446	81.53	2105	84.94	370	92.97	NC_022134.1	Wang et al., 2013
Timelaea maculata	15178	81.03	-0.036	-0.195	11176	3726	79.65	1454	81.57	2109	85.39	382	93.19	NC_021090.1	Wang et al., unpublished
Biblidinae															
Ariadne ariadne	15179	80.07	-0.028	-0.216	11194	3732	78.75	1461	80.83	2119	84.00	319	92.79	NC_026069.1	Shi et al., 2015
Hamadryas epinome	15207	81.50	-0.031	-0.189	11202	3734	80.35	1451	81.25	2129	84.92	337	93.17	NC_025551.1	Cally et al., 2014
Calinaginae															
Calinaga davidis	15267	80.45	-0.044	-0.201	11247	3749	78.94	1457	81.67	2110	84.60	389	92.03	NC_015480.1	Xia Jing, 2011
Charaxinae															
Polyura nepenthes	15333	80.79	-0.033	-0.199	11212	3738	79.45	1463	81.95	2161	84.77	437	89.47	NC_026073.1	Shi et al., 2015
Danainae															
Danaus chrysippus	15236	80.37	-0.031	-0.238	11176	3726	78.81	1451	81.60	2122	84.35	418	94.98	NC_024532.1	Gan et al., 2014
Heliconiinae															
Heliconius hecale	15338	81.63	-0.033	-0.188	11172	3724	80.15	1457	82.43	2141	85.15	469	95.52	KM068091.1	Shen & Wang, 2014
Libytheinae															
Libythea celtis	15164	81.21	-0.017	-0.184	11197	3733	80.00	1459	80.94	2110	84.98	328	96.34	NC_016724.1	Hao et al., 2013a
Limenitidinae															
Athyma sulpitia	15268	81.95	-0.047	-0.176	11229	3743	80.70	1446	81.53	2098	85.08	349	94.55	NC_017744.1	Tian et al., 2012
Pseudergolinae															
Dichorragia nesimachus	15355	81.67	-0.029	-0.188	11178	3726	80.21	1452	82.37	2113	85.28	55	92.73	NC_024409.1	Wu et al., 2014

Satyrinae															
Triphysa phryne	15143	78.83	-0.036	-0.211	11185	3729	77.31	1453	80.66	2118	84.51	316	83.54	NC_024551.1	Zhang et al., 2014

The parameters I and G were estimated in PAUP* 4.0b8/4.0d78, for both the barcode sequences and the complete mitochondrial genomes for maximum likelihood phylogenetic analyses with GTR model settings using 10 heuristic search replicates with random sequence and tree bisection and reconnection branch swapping on best trees only, and retaining all best trees (otherwise all PAUP default settings were used) as recommended by the PAUP user manual (Swofford, 2002). For the whole mitochondrial genome data set, the proportion of invariable sites (I) was found to be 0.372773and the estimated gamma (G) was 0.765056. For the first and second codon positions of protein coding genes data set, the proportion of invariable sites (I) was found to be 0.389029 and the estimated gamma (G) was 0.818619. These values were then used for more comprehensive heuristic and bootstrap searches in PAUP. The sequences were analyzed phylogenetically with the maximum likelihood (GTR+I+G model) and parsimony algorithms first with 1 million heuristic searches with random starting trees and tree bisection and reconnection branch swapping, followed by bootstrap analysis with 1 million quick addition search replicates (otherwise, default PAUP settings were used) (Swofford, 2002).

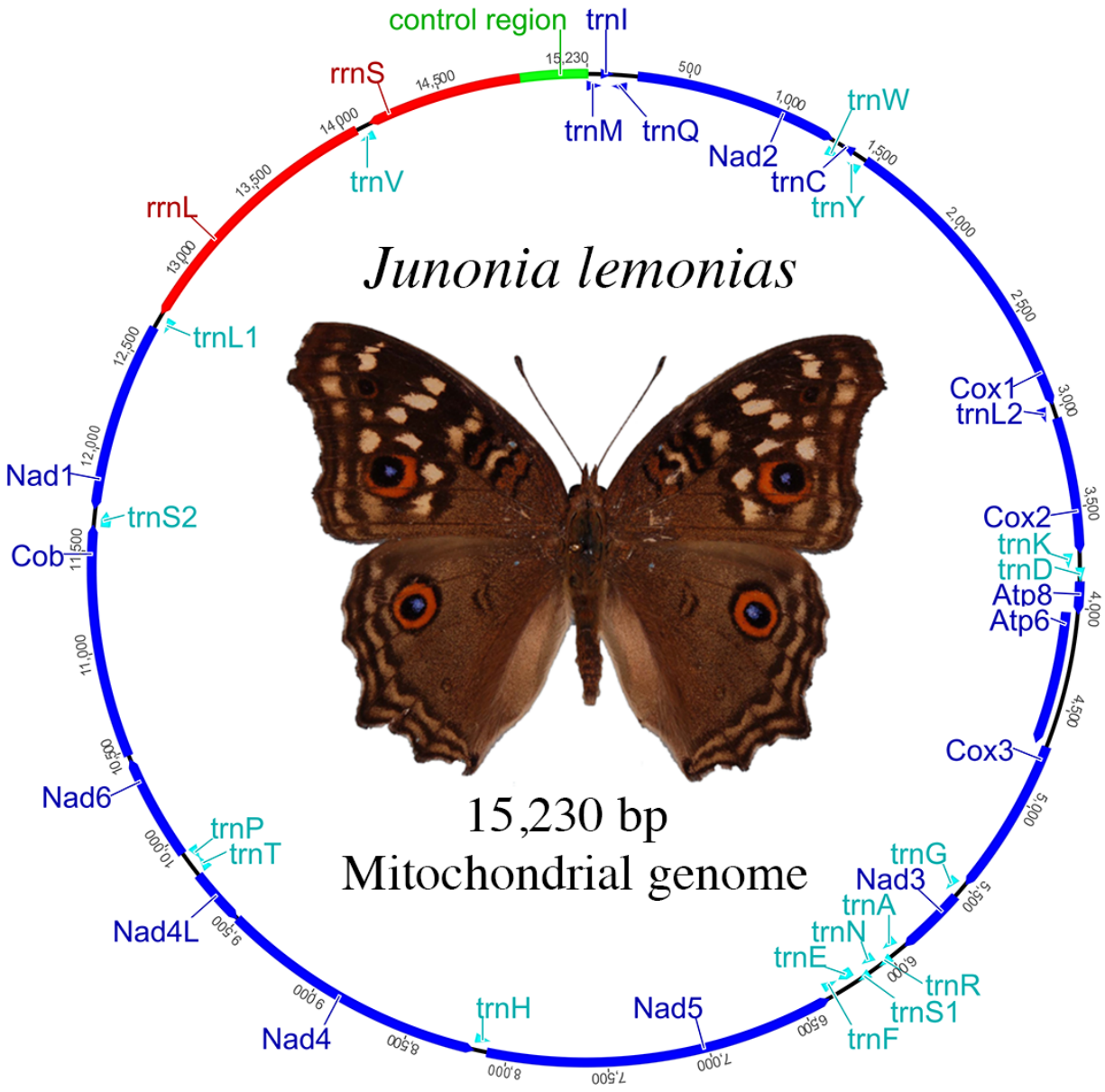
Results and discussion

Next generation sequencing output and mitochondrial genome organization

A total of 11,447,872 Illumina MiSeq reads were produced from the *J. lemonias* indexed library. Of those reads, 46,304 sequence reads were aligned to assemble the *J. lemonias* mitochondrial genome. The mean base coverage of the mitochondrial genome was 634.8-fold with a maximum coverage of 1,385-fold and a minimum 169-fold coverage. The complete *J. lemonias* mitochondrial genome is 15,230 bp and has been assigned GenBank accession number KP941756 (Fig. 2-1).

The *J. lemonias* mitochondrial genome is intermediate in size between the previously sequenced mitochondrial genomes of *J. orithya* (15,214 bp) (Shi et al., 2013) and *Junonia almana* (15,256 bp) (Wu et al., 2014). Like other butterfly mitochondrial genomes, *J. lemonias* contains 13 protein-coding genes, 22 transfer RNA genes, and 2 ribosomal RNA genes (Table 2-2). The order and arrangement of the genes is identical to that of other *Junonia* species (Shi et al., 2013, Wu et al., 2014) and is typical of other Ditrysian Lepidoptera (Cao et al., 2012). The total length of mitochondrial protein-coding sequence in *J. lemonias* is 11,190 bp. Additionally, the *J. lemonias* mitochondrial genome contains a control region sequence consisting of 333 bp. There are 11 regions (35 bp overall) in the mitochondrion where adjacent genes overlap. Other than the control region, there are 15 intergenic regions consisting of 151 bp total. The largest of these intergenic regions is between trnQ and *nad2* and is 51 bp. The mitogenome is 80.40% AT content and 19.60% GC content. The *J. lemonias* mitogenome AT content is identical to the published *J. orithya* mitogenome (Shi et al., 2013) and is similar to the AT content of *J. almana* (80.26%) (Wu et al., 2014). The nucleotide composition of the strand with the majority of the protein-coding sequences (which is also the light, pyrimidine-rich mitochondrial strand) in *J. lemonias* is skewed toward T (AT skew value = -0.009), and C (GC skew value = -0.224). This is similar to calculated values for the mitochondrial genomes of *J. almana* (AT skew=-0.01; GC skew=-0.227) and *J. orithya* (AT skew=-0.01; GC skew=-0.179) (Table 2-1).

Fig. 2-1. Circular map of the mitochondrial genome of *J. lemonias*, the lemon pansy butterfly. Specimen depicted is in the collection of the Yale Peabody Museum of Natural History (catalog number YPM ENT 412062), collected Victoria Peak, Hong Kong, 6 September 1966 by R. W Elliot. Mitochondrial genome annotation depiction was made with Geneious 8.1.3. Gene abbreviations in the figure are: *atp6 (ATP synthase subunit 6), atp8 (ATP synthase subunit 8), cob (cytochrome b), cox1 (Cytochrome C oxidase subunit I), cox2 (Cytochrome C oxidase subunit II), cox3 (Cytochrome C oxidase subunit III), nad1 (NADH dehydrogenase subunit 1), nad2 (NADH dehydrogenase subunit 2), nad3 (NADH dehydrogenase subunit 3), nad4 (NADH dehydrogenase subunit 4), nad4L (NADH dehydrogenase subunit 4L), nad5 (NADH dehydrogenase subunit 5), nad6 (NADH dehydrogenase subunit 6).* tRNA abbreviations in figure: *trnM (tRNA Methionine), trnQ (tRNA Glutamine), trnI (tRNA Isoleucine), trnK (tRNA Lysine), trnD (tRNA Aspartic acid), trnG (tRNA Glycine), trnL1 (tRNA Leucine 1), trnL2 (tRNA Leucine2), trnA (tRNA Alanine), trnN (tRNA Asparagine), trnE (tRNA Glutamic acid), trnF (tRNA Phenylalanine), trnR (tRNA Arginine), trnS1 (tRNA Serine 1), trnH (tRNA Histidine), trnT (tRNA Threonine), trnP (tRNA Proline), trnS2 (tRNA Serine 2), trnV*. rRNA abbreviations in figure: *rrnL (ribosomal RNA: large subunit), rrnS (ribosomal RNA: small subunit)*.



The *cox1* gene of *J. lemonias* has an aberrant start codon (CGA) (which is typical of insect mitochondrial genomes (Cameron, 2014) and there are two protein-coding genes (*cox1, cox2*) that have aberrant single-nucleotide (T) stop codons as has been reported previously in other *Junonia* mitochondrial genomes (Shi et al., 2013, Wu et al., 2014). Other aberrant single-nucleotide (T) stop codons (*nad1* and *nad4*) previously reported in other *Junonia* mitochondrial genomes (Shi et al., 2013, Wu et al., 2014) appear to be annotation errors. Computational

modeling of tRNA structures with RNAstructure (Reuters 2010) shows that all but one tRNA conform to the standard tRNA cloverleaf secondary structure. The remaining trnS (AGN) has a loop in place of a dihydrouridine arm, as also seen in *J. orithya* (Shi et al., 2013), other butterfly species, and almost all arthropods (Xu et al., 2009, Qin et al., 2012, Wang et al., 2015). The two rRNAs in *J. lemonias* both have very high AT content, with rrnL consisting of 84.01% AT and rrnS consisting of 84.88% AT.

Table 2-2. Arrangement of the mitochondrial genome of *J. lemonias*.

Gene	Direction	Position	Size (bp)	Intergenic length	Start Codon	Stop Codon
trnM	F	1–68	68			
trnaI	F	67–133	67	-2		
trnQ	R	131-199	69	-3		
nad2	F	251–1264	1014	51	ATT	TAA
trnW	R	1263–1329	67	-2		
trnC	R	1322–1383	62	-8		
trnY	R	1389–1453	65	4		
cox1	F	1456–2986	1531	2	CGA	T
trnL2	F	2987–3054	68	0		
cox2	F	3055–3730	676	0	ATG	T
trnK	F	3731–3801	71	0		
trnD	F	3801–3866	66	-1		
atp8	F	3867–4031	165	0	ATT	TAA
atp6	F	4025–4702	678	-7	ATG	TAA
cox3	F	4702–5490	789	-1	ATG	TAA
trnG	F	5493–5560	68	2		
nad3	F	5561–5914	354	0	ATC	TAA
trnA	F	5933–6002	70	18		
trnR	F	6002–6066	65	-1		
trnN	F	6067–6132	66	0		
trnS1	F	6131–6191	61	-2		
trnE	F	6203–6275	73	11		
trnF	R	6296–6361	66	20		
nad5	R	6365–8098	1734	3	ATT	TAA
trnH	R	8099–8167	69	0		
nad4	R	8162–9506	1340	-6	ATG	TAA
nad4l	R	9507–9794	288	0	ATG	TAA
trnT	F	9797–9860	64	2		
trnP	R	9861–9925	65	0		
nad6	F	9928–10455	528	2	ATT	TAA
Cob	F	10478–11629	1152	22	ATG	TAA
trnS2	F	11628–11697	70	-2		

nad1	R	11707–12652	938	9	ATG	TAA
trnL1	R	12654–12724	71	1		
rrnL	R	12728–14058	1331	3		
trnV	R	14059–14122	64	0		
rrnS	R	14124–14897	774	1		
control region		14898–15230	333	0		

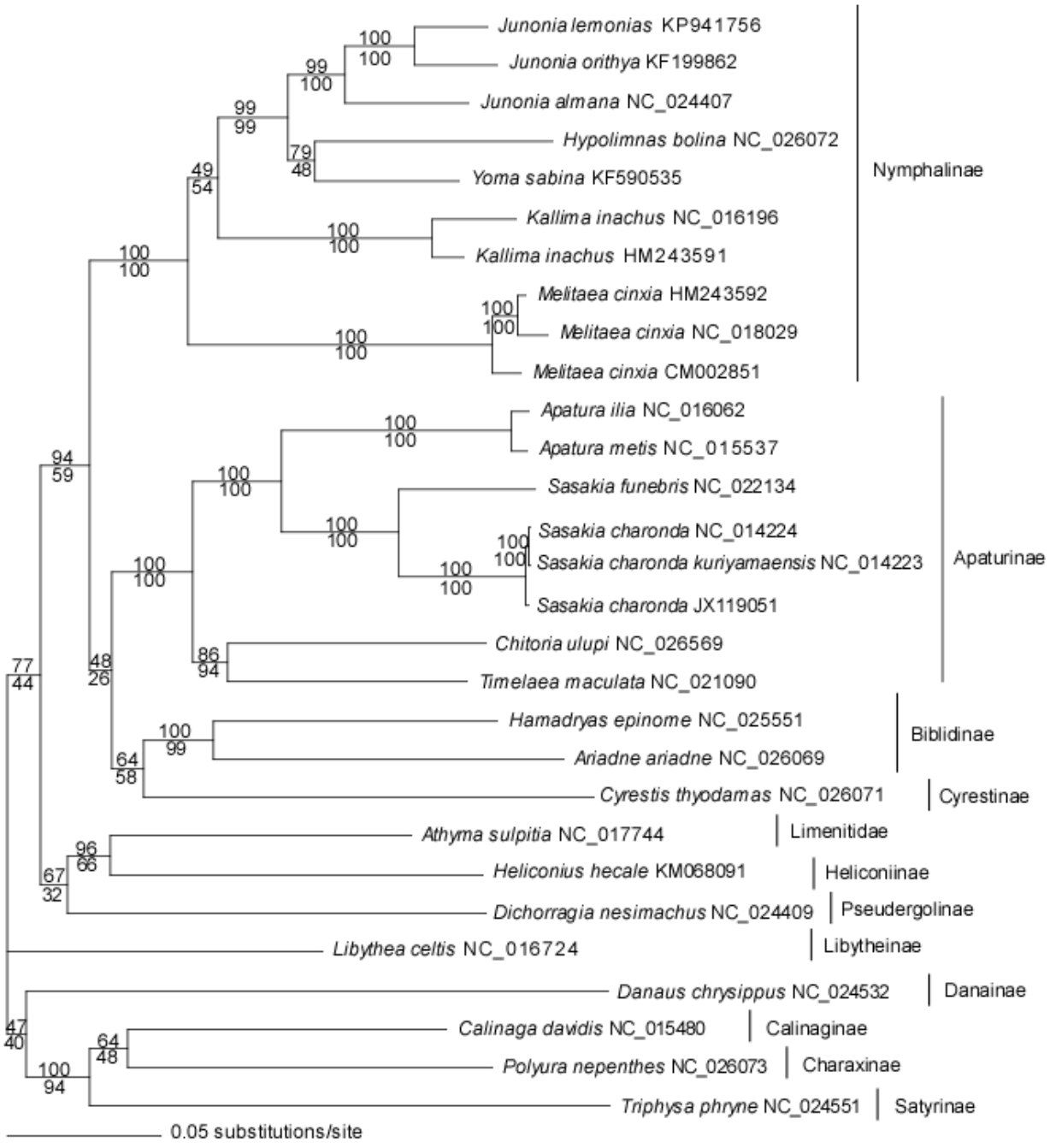
The 333 bp control region has the highest AT content (94.29%) in the mitochondrial genome of *J. lemonias*. The control region includes bases 14,898 to 15,230. This AT-rich region lies between rrnS and trnM. Like *J. orithya* and other Lepidopteran mitogenomes (Wang et al., 2015), the control region includes an ATAGA motif (Cameron & Whiting, 2008), although in *J. lemonias* it is followed by a 20 bp poly-T stretch as opposed to the 18 bp poly-T stretch found in *J. orithya* (Shi et al., 2013). *J. almana* also has this ATAGA motif with a 19 bp poly-T stretch, but the motif has an extra terminal A nucleotide so that the sequence is ATAGAA. Additionally, there are two microsatellite-like elements in the control region: one $(TA)_8$ and one $(TA)_9$. The length of the control region is comparable to the average (334.6 bp) of the other nymphalid mitochondrial genomes used for our phylogenetic analysis.

Across the nymphalid species that we have analyzed, some interesting comparisons can be made (Table 2-1). The overall length of tRNA sequences in *J. lemonias* was actually the largest (1475 bp), with the shortest tRNA lengths (1446 bp) being found in *Sasakia funebris* and *Athyma sulpitia.* Within the mitochondrial genomes available from within the Nymphalinae, *J. lemonias* had the second highest percentage of AT content (81.00%), with *Y. sabina* having the highest (81.11%).

Clustal W2 alignment of the mitogenome sequences revealed several indels in intergenic regions but none of these indels created ambiguities in the alignment. Maximum likelihood and

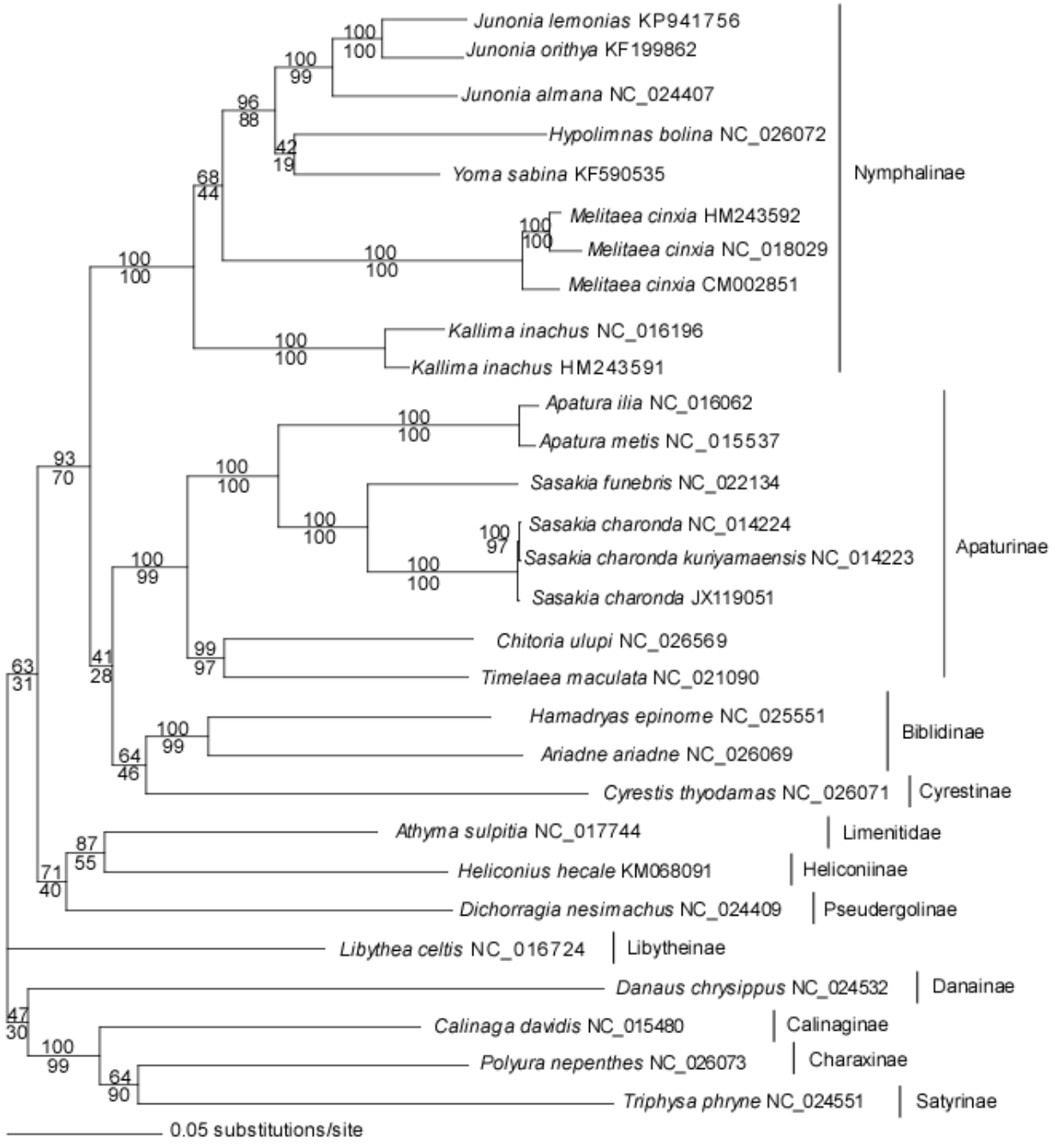
parsimony analysis of the full-length mitogenome sequences produced a single optimal phylogenetic tree (Fig. 2-2). Within the Nymphalinae, the monophyly of the Nymphalinae and the genus *Junonia* have 99-100% maximum likelihood and parsimony bootstrap support. However, the sister relationship of *H. bolina* and *Y. sabina* has much weaker bootstrap support (79% maximum likelihood, 48% parsimony). Additionally, the placement of *K. inachus* mitochondrial genomes to the rest of the Nymphalinae has modest (49% maximum likelihood, 54% parsimony) maximum likelihood and parsimony bootstrap support. The relationships of the genera within the Apaturinae are all well supported. The relationship of the Apaturinae and the clade containing Biblidinae and Cyrestinae has less bootstrap support (48% maximum likelihood, 26% parsimony). Although Limenitidae and Heliconiinae have moderate bootstrap support as sister clades (96% maximum likelihood, 66% parsimony), the support for their grouping with as sister to Pseudergolinae has weak bootstrap support (67% maximum likelihood, 32% parsimony). Our tree shows that *Dichorragia nesimachus* and *Cyrestis thyodamas* do not form a monophyletic group and thus agrees with earlier workers who removed *D. nesimachus* from nymphalid subfamily Cyrestinae and reassigned it to subfamily Pseudergolinae (Wahlberg et al., 2005). Overall, the bootstrap support for basal nodes between nymphalid subfamilies is generally low, so phylogenetic analyses of complete mitochondrial genome sequences may be most useful for understanding species relationships within subfamilies.

Fig. 2-2. Maximum likelihood tree of complete mitochondrial genome sequences. Maximum likelihood bootstrap values are above each node, with parsimony bootstrap values below each node.



Additional maximum likelihood and parsimony phylogenetic trees derived from only mitochondrial protein-coding sequences with third codon positions excluded (Fig. 2-3) were generated in PAUP (Swofford, 2002). Like the trees derived from complete mitochondrial genomes, the subfamily Nymphalinae and the genus *Junonia* continues to have strong bootstrap support. As in the analyses of the complete mitochondrial sequences, analysis of first and second codons in protein-coding sequences yields a tree where Nymphalinae is sister to a clade containing the Apaturinae, Biblidinae, and Cyrestinae, but the clade has a bootstrap value of only 41% maximum likelihood and 28% parsimony bootstrap (Fig. 2-3). The position of *Melitaea cinxia* to the rest of the Nymphalinae (excluding *Kallima inachus*) has modest bootstrap support 68% maximum likelihood, 44% parsiomony). The relationships within the Apaturinae have consistently strong bootstrap support.

Fig. 2-3. Maximum likelihood tree of coding sequences, with third codon positions excluded, from mitochondrial genomes. Maximum likelihood bootstrap values are above each node, with parsimony bootstrap values below each node.



By removing the areas of the mitochondrial genome most likely to cause noise (limiting analysis to only sequences from protein coding genes with third codon positions removed) in phylogenetic analysis, bootstrap support increased for some nodes. For example, the sister clade relationship of the Charaxinae and Satyrinae had 53% maximum likelihood and 52% parsimony bootstrap support based on whole mitochondrial genomes and 93% maximum likelihood and 90% parsimony bootstrap support based on analysis of first and second codon positions. At the same time, restricting the analysis to first and second codon positions also yielded a number of phylogenetic relationships among the nymphalid subfamilies that were not strongly supported. For example, the sister relationship between the Biblidinae and Cyrestinae had lower bootstrap values in this analysis (53% maximum likelihood and 46% parsimony bootstrap values) than in the analysis of complete mitochondrial genomes (73% maximum likelihood and 58% parsimony bootstrap values).

Other previously published phylogenetic reconstructions have placed the Cyrestinae (Wahlberg & Wheat, 2008, Wahlberg et al., 2009) or the Apaturinae (Wu et al., 2014) as the sister clade to the Nymphalinae, all generally with low posterior probabilities or bootstrap support. Our analyses of both whole mitochondrial genomes (Fig. 2-2) and first and second positions of mitochondrial coding sequences (Fig. 2-3) suggest an alternative, although also with low bootstrap support. Our data suggest that the Nymphalinae may have a sister relationship to a clade containing the subfamilies Apaturinae, Biblidinae, and Cyrestinae. The node supporting the close relationship of these 3 subfamilies to the Nymphalinae has 81% maximum likelihood and 59% parsimony bootstrap support in analyses of whole mitochondrial genomes and 83% maximum likelihood (Fig. 2-2) and 70% parsimony bootstrap support in analyses of first and second codon positions in mitochondrial protein-coding genes (Fig. 2-3). Completely resolving

the relationships of the subfamilies within the Nymphalidae with strong bootstrap support will likely require molecular phylogenetic data sets that include either far more extensive taxon sampling of Nymphalid mitochondrial genomes to reduce branch lengths, better identify homoplasy, and improve tree construction; sequence data from nuclear genes to increase the number of informative characters for phylogenetic reconstruction; or a combination of the two.

Finally, while it is now clear that the genus *Junonia* originated and diversified in Africa, followed by an emigration event to southeast Asia where a secondary radiation took place (Kodandaramaiah & Wahlberg, 2007, Gaikwad et al., 2012), there is still some ambiguity about the colonization history of the New World by this genus (Gemmell et al., 2014, Gemmell & Marcus, 2015) and hypotheses for the colonization of the New World from either Africa or Asia have been proposed (Eliot, 1946, Forbes, 1947). Thus far, only mitochondrial genomes from Asian *Junonia* species have been sequenced, so it would be highly beneficial to begin accumulating data from African and New World species as well. Phylogenomic approaches involving the *Junonia* mitochondrial genome have great potential to clarify the phylogenetic history of this group, permitting studies of the evolution of seasonal polyphenism, color pattern evolution and development, and the colonization of novel larval host plants in this genus, as well as to permit studies of the molecular evolution of the mitochondrial genome itself.

Acknowledgements

We thank Robert Dowell (California Department of Food and Agriculture) for his field collection of *Junonia lemonias*. We thank Larry Gall and the Yale Peabody Museum Entomology Collection for permission to use a photograph of a *J. lemonias* specimen from their

holdings. Karl Gordon, Tom Walsh, and Weetek Tay very generously shared advice about their experiences with next generation sequencing of mitochondrial genomes during the design stage of this project. We sincerely thank Deborah Tsuyuki and the Next Generation Sequencing Platform at the Manitoba Institute of Child Health for their assistance. We thank members of the Marcus lab for their help and camaraderie. Si Hyeock Lee and two anonymous reviewers provided helpful suggestions on an earlier draft of the manuscript. This research received support from a University of Manitoba Graduate Fellowship (to B.S.M.) and an NSERC Discovery Grant (RGPIN386337-2011), a Canada Foundation for Innovation Award (212382), and the Canada Research Chair program (950–212382) (to J.M.M.).

Literature Cited

Ahmad, V.U., Ali, A., Ali, Z., Baqai, F.T. and Naz Zafar, F., 1998. Cycloartane triterpene glucosides from *Corchorus depressus*. Phytochem. 49, 829-834.

Beaudette, K., Hughes, T.M. and Marcus, J.M., 2014. Improved injection needles facilitate germ-line transformation of the buckeye butterfly *Junonia coenia*. Biotechniques 56, 142-144. doi: 10.2144/000114147.

Borchers, T.E. and Marcus, J.M., 2014. Genetic population structure of buckeye butterflies (*Junonia*) from Argentina. Syst. Ent. 39, 242-255. doi: 10.1111/syen.12053.

Bowers, M.D., 1984. Iridoid Glycosides and Host-Plant Specificity in Larvae of the Buckeye Butterfly, *Junonia coenia* (Nymphalidae). J. Chem. Ecol. 10, 1567-1577.

Brévignon, L. and Brévignon, C., 2012. Nouvelles observations sur le genre *Junonia* en Guyane Française. (Lepidoptera: Nymphalidae) (Seconde partie). Lépidoptères de Guyane 7, 8-35.

Brown, F.M. and Heineman, B., 1972. Jamaica and its Butterflies., E. W. Classey, Ltd., London.

Cally, S., Lhuillier, E., Iribar, A., Garzón-Orduña, I., Coissac, E. and Murienne, J., 2014. Shotgun assembly of the complete mitochondrial genome of the neotropical cracker butterfly *Hamadryas epinome*. Mitochondrial DNA 0, 1-3 doi:10.3109/19401736.2014.971262.

Camara, M.D., 1997. A recent host range expansion in *Junonia coenia* Huebner (Nymphalidae): Oviposition preference, survival, growth, and chemical defense. Evolution 51, 873-884.

Cameron, S.L., 2014. How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. Syst. Entomol. 39, 400-411.

Cameron, S.L. and Whiting, M.F., 2008. The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. Gene 408, 112-123.

Cao, Y.Q., Ma, C., Chen, J.Y. and Yang, D.R., 2012. The complete mitochondrial genomes of two ghost moths, *Thitarodes renzhiensis* and *Thitarodes yunnanensis*: the ancestral gene arrangement in Lepidoptera. BMC Genomics 13, 276.

Carroll, S.B., Gates, J., Keys, D.N., Paddock, S.W., Panganiban, G.E.F., Selegue, J.E. and Williams, J.A., 1994. Pattern formation and eyespot determination in butterfly wings. Science 265, 109-114.

Chen, M., Li-Li, T., Shi, Q.-H., Cao, T.-W. and Hao, J.-S., 2012. Complete mitogenome of the Lesser Purple Emperor *Apatura ilia* (Lepidoptera: Nymphalidae: Apaturinae) and comparison with other nymphalid butterflies. Zool. Res. pp. 191-201.

Chou, I., (Ed.) 1994. Monographia Rhopalocerum Sinensium, Henan Scientific and Technological Publishing House, China.

Darriba, D., Taboada, G.L., Doallo, R. and Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Meth. 9, 772-772.

Dhungel, B., Ohno, Y., Matayoshi, R. and Otaki, J.M., 2013. Baculovirus-mediated gene transfer in butterfly wings *in vivo*: an efficient expression system with an anti-gp64 antibody. BMC Biotechnol. 13, 27. doi: 10.1186/1472-6750-13-27.

Dhungel, B. and Otaki, J.M., 2013. Larval temperature experience determines sensitivity to cold-shock-induced wing color pattern changes in the blue pansy butterfly Junonia orithya. J. Thermal Biol. 38, 427-433.

Ekpo, M.A. and Etim, P.C., 2009. Antimicrobial activity of ethanolic and aqueous extracts of *Sida acuta* on microorganisms from skin infections. J. Med. Plants Res. 3, 621-624.

Eliot, N., 1946. Continental drift and *Precis lavinia.* Entomologist 79, 225-228.

ElSohly, M.A. and Slade, D., 2005. Chemical constituents of marijuana: The complex mixture of natural cannabinoids. Life Sciences 78, 539-548 doi:10.1016/j.lfs.2005.09.011.

Emmel, T.C. and Emmel, J.F., 1973. The butterflies of Southern California. Nat. Hist. Mus. Los Angeles Co. Sci Ser. 26, 1-148.

EndMemo, 2015. DNA/RNA GC Content Calculator. EndMemo.

Forbes, W.T.M., 1947. Buckeyes and Wegener. Entomologist 80, 56-58.

Gaikwad, S.S., Ghate, H.V., Ghaskadbi, S.S., Patole, M.S. and Shouche, Y.S., 2012. DNA barcoding of nymphalid butterflies (Nymphalidae: Lepidoptera) from Western Ghats of India. Mol. Biol. Rep. 39, 2375-2383.

Gan, S.-S., Sun, X.-Y., Gai, Y.-H. and Hao, J.-S., 2014. The complete mitochondrial genome of *Danaus chrysippus* (Lepidoptera: Nymphalidae: Danainae). Mitochondrial DNA 26, 819-820.

Gemmell, A.P., Borchers, T.E. and Marcus, J.M., 2014. Genetic Population Structure of Buckeye Butterflies (*Junonia*) from French Guiana, Martinique, and Guadeloupe. Psyche 2014, 1-21. doi: 10.1155/2014/897596.

Gemmell, A.P. and Marcus, J.M., 2015. A tale of two haplotype groups: The origin and distribution of divergent New World *Junonia COI* haplotypes. Syst. Ent. 40, 532-546. doi: 10.1111/syen.12120.

Gillett, C.P.D.T., Crampton-Platt, A., Timmermans, M.J.T.N., Jordal, B.H., Emerson, B.C. and Vogler, A.P., 2014. Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). Mol. Biol. Evol. 31, 2223-2237.

Gu, M. and Chen, P.H., 1997. Butterflies in Hainan Island, China Forestry Publishing House, Beijing.

Hao, J., Sun, M., #x1f5, Shi, Q., Sun, X., Shao, L. and Yang, Q., 2013. Complete Mitogenomes of *Euploea mulciber* (Nymphalidae: Danainae) and *Libythea celtis* (Nymphalidae: Libytheinae) and Their Phylogenetic Implications. ISRN Genomics 2013, 14.

Huelsenbeck, J.P. and Rannala, B., 1997. Phylogenetic methods come of age: Testing hypotheses in an evolutionary context. Science 276, 227-232.

Igarishi, S. and Fukada, H., 1997. The Life Histories of Asian Butterflies, Volume 1, Tokai University Press.

Jaiswal, S.K., Dubey, M.K., Das, S. and Rao, C.V., 2014. Gastroprotective effect of the iridoid fraction from *Barleria prionitis* leaves on experimentally-induced gastric ulceration. Chin. J. Nat. Med. 12, 738-744 doi: 10.1016/S1875-5364(14)60113-8.

Jensen, H.F.W., Jensen, S.R. and Neilsen, B.J., 1988. Chemotaxonomy of the Acanthaceae. Iridoids and quaternary amines. Phytochem. 27, 2581-2589.

Jensen, S.R. and Neilsen, B.J., 1985. Hygrophiloside, an iridoid glusoside from *Hygrophila difformis* (Acanthaceae). Phytochem. 24, 602-603.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647-1649.

Kim, K.H., Park, Y.J., Chung, K.H., Yip, M.L.R., Clardy, J., Senger, D. and Cao, S., 2015. Iridoid Glycosides from *Barleria lupulina*. J. Nat. Prod. 78, 320–324 DOI: 10.1021/np500791a.

Knerl, A. and Bowers, M.D., 2013. Incorporation of an introduced weed into the diet of a native butterfly: Consequences for preference, performance and chemical defense. J. Chem. Ecol. 39, 1313-1321. doi: 10.1007/s10886-013-0355-3.

Kodandaramaiah, U., 2009. Eyespot evolution: Phylogenetic insights from *Junonia* and related butterfly genera (Nymphalidae: Junoniini). Evol. Dev. 11, 489-497.

Kodandaramaiah, U. and Wahlberg, N., 2007. Out-of-Africa origin and dispersal-mediated diversification of the butterfly genus *Junonia* (Nymphalidae: Nymphalinae). J. Evol. Biol. 20, 2181-2191.

Kunte, K., 2000. Butterflies of Peninsular India, Universities Press, Hyderabad, India.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948.

Linnaeus, 1758. Systema Naturae per Regna Tria Naturae, Secundum Clases, Ordines, Genera, Species, cum Characteribus, Differentiis, Symonymis, Locis., Syst. Nat. .

Lobry, J.R., 1996. Asymmetric substitution patterns in the two DNA strands of bacteria. Mol. Biol. Evol. 13, 660-665.

Martin, A. and Reed, R.D., 2014. Wnt signaling underlies evolution and development of the butterfly wing pattern symmetry systems. Dev. Biol. 395, 367-378.

McDade, L.A., Daniel, T.F. and Kiel, C.A., 2008. Toward a comprehensive understanding of phylogenetic relationships among lineages of Acanthaceae s.l. (Lamiales). Am. J. Bot. 95, 1136-1152 doi: 10.3732/ajb.0800096.

McLeod, L., 1984. Seasonal polyphenism in African *Precis* butterflies. The Biology of Butterflies. . (ed. R.I. Vane-Wrights), pp. 313-315. Academic Press, London.

Nijhout, H.F., 2010. Molecular and Physiological basis of color pattern formation. Adv. Insect Physiol. 38, 219-265.
Nijhout, H.F. and Grunert, L.W., 2002. Bombyxin is a growth factor for wing imaginal disks in Lepidoptera. Proc. Nat. Acad. Sci. USA 99, 15446-15450.
Otaki, J.M., 2012. Structural analysis of eyespots: dynamics of morphogenic signals that govern elemental positions in butterfly wings. BMC Syst. Biol. 6, 17. doi: 10.1186/1752-0509-6-17.
Pereyra, P.C. and Bowers, M.D., 1988. Iridoid glycosides as oviposition stimulants for the buckeye butterfly, *Junonia coenia* (Nymphalidae). J. Chem. Ecol. 14, 917-928.
Pfeiler, E., Johnson, S. and Markow, T.A., 2012. DNA barcodes and insights into the relationships and systematics of buckeye butterflies (Nymphalidae: Nymphalinae: *Junonia*) from the Americas. J. Lepid. Soc. 66, 185-198.
Qin, X.-M., Guan, Q.-X., Zeng, D.-L., Qin, F. and Li, H.-M., 2012. Complete mitochondrial genome of *Kallima inachus* (Lepidoptera: Nymphalidae: Nymphalinae): Comparison of *K. inachus* and *Argynnis hyperbius*. Mitochondrial DNA 23, 318-320 doi: 10.3109/19401736.2012.684093.
Rane, N.S. and Ranade, S.P., 2004. Butterflies of the Tamhini-Dongarwadi Area, Mulshi, Maharashtra. Zoos' Print Journal 19, 1411-1413.
Richards, L.A., Lampert, E.C., Bowers, M.D., Dodson, C.D., Smilanich, A.M. and Dyer, L.A., 2012. Synergistic effects of iridoid glycosides on the survival, development and immune response of a specialist caterpillar, *Junonia coenia* (Nymphalidae). J. Chem. Ecol. 38, 1276-1284.
Rutowski, R.L., 1991. Temporal and spatial overlap in the mate-locating behavior of two species of *Junonia* (Nymphalidae). J. Res. Lep. 30, 267-271.
Shen, Q.-Q. and Wang, L., 2014. The complete mitochondrial genome sequence of *Heliconius hecale* (Insecta: Lepidoptera: Nymphalidae). Mitochondrial DNA 0, 1-2 DOI: 10.3109/19401736.2014.945561.
Shi, Q., Huang, D., Wang, Y. and Hao, J., 2013. The complete mitochondrial genome of Blue Pansy, *Junonia orithya* (Lepidoptera: Nymphalidae: Nymphalinae). Mitochondrial DNA 0, 1-2. doi: 10.3109/19401736.19402013.19823182.
Smith, K.C., 1991. The effects of temperature and daylength on the *Rosa* polyphenism in the buckeye butterfly, *Precis coenia* (Lepidoptera: Nymphalidae). J. Res. Lep. 30, 225-236.
Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts, USA.
Tian, L.L., Sun, X.Y., Chen, M., Gai, Y.H., Hao, J.S. and Yang, Q., 2012. Complete mitochondrial genome of the Five-dot Sergeant *Parathyma sulpitia* (Nymphalidae: Limenitidinae) and its phylogenetic implications. Zool. Res. 33, 133-143. doi: 10.3724/SP.J.1141.2012.02133.
Turner, C.E., ElSohly, M.A. and Boeren, E.G., 1980. Constituents of *Cannabis sativa* L. XVII. A review of the natural constituents. J. Nat. Prod. 43, 169-234.
Turner, T.W. and Parnell, J.R., 1985. The identification of two species of *Junonia* Hübner (Lepidoptera: Nymphalidae): *J. evarete* and *J. genoveva* in Jamaica. J. Res. Lep. 24, 142-153.
Vane-Wright, R.I. and Tennent, W.J., 2011. Colour and size variation in *Junonia villida* (Lepidoptera, Nymphalidae): subspecies or phenotypic plasticity? Syst. and Biodivers. 9, 289-305.

Wahlberg, N., Brower, A.V.Z. and Nylin, S., 2005. Phylogenetic relationships and historical biogeography of tribes and genera in the subfamily Nymphalinae (Lepidoptera: Nymphalidae). Biol. J. Linn. Soc. 86, 227-251.

Wahlberg, N., Weingartner, E., Warren, A. and Nylin, S., 2009. Timing major conflict between mitochondrial and nuclear genes in species relationships of *Polygonia* butterflies (Nymphalidae: Nymphalini). BMC Evol. Biol. 9, 92.

Wahlberg, N. and Wheat, C., 2008. Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers. Syst. Biol. 57, 231-242.

Wang, J., Nie, X., Cao, T., Zhang, M., Guo, Y., Ma, E. and Zhang, X., 2012. Analysis of complete mitochondrial genome of *Sasakia charonda coreana* (Lepidoptera, nymphalida). Acta Zootaxonomica Sinica 37, 1-9.

Wang, J.P., Cao, T.W., Xuan, S.B., Wang, H., Zhang, M. and Ma, E.b., 2013. The complete mitochondrial genome of *Sasakia funebris* (Leech) (Lepidoptera: Nymphalidae) and comparison with other Apaturinae insects. Gene 526, 336-343.

Wang, Y.-L., Chen, Y.-H., Xia, C.-C., Xia, X.-Q., Tao, R.-S. and Hao, J.-S., 2015. The complete mitochondrial genome of the Common Red Apollo, *Parnassius epaphus* (Lepidoptera: Papilionidae: Parnassiinae). J. Asia Pac. Ent. 18, 239-248 http://dx.doi.org/10.1016/j.aspen.2015.02.002.

Wernersson, R., 2005. FeatureExtract—extraction of sequence annotation made easy. Nucleic Acids Res 33, W567-W569. doi: 10.1093/nar/gki388.

Wu, L.-W., Lin, L.-H., Lees, D. and Hsu, Y.-F., 2014. Mitogenomic sequences effectively recover relationships within brush-footed butterflies (Lepidoptera: Nymphalidae). BMC Genomics 15, 468 doi:10.1186/1471-2164-15-468.

Xia Jing, H.J., Zhu Guo-Ping, Zhu Chao-Dong, Hao Jia-Sheng, 2011. Sequencing and analysis of the complete mitochondrial genome of *Calinaga davidis* Oberthür (Lepidoptera: Nymphalidae). Acta Entomologica Sinica 54, 555-565.

Xu, C., Chen, H., Wang, R. and Xu, C., 2009. The mitochondrial genome of the Glanville fritillary butterfly, *Melitaea cinxia* (Lepidoptera: Nymphalidae). Genbank accession GQ398377.

Yu, H.W., Li, B.G., Li, G.Y., Li, C.S., Fang, D.M. and Zhang, G.L., 2007. Abietane lactones and iridoids from *Goldfussia yunnanensis*. Chem. Pharm. Bull. (Tokyo) 55, 1744-1747.

Zhang, M., Nie, X., Cao, T., Wang, J., Li, T., Zhang, X., Guo, Y., Ma, E. and Zhong, Y., 2012. The complete mitochondrial genome of the butterfly *Apatura metis* (Lepidoptera: Nymphalidae). Mol. Biol. Rep. 39, 6529-6536.

Zhang, W., Gan, S., Zuo, N., Chen, C., Wang, Y. and Hao, J., 2014. The complete mitochondrial genome of Triphysa phryne (Lepidoptera: Nymphalidae: Satyrinae). Mitochondrial DNA, 1-2.

Chapter 3: Complete mitochondrial genome sequences reveal that the New World Junonia (Lepidoptera: Nymphalidae) are not monophyletic

Abstract

DNA *cox1* barcode sequences are a popular tool for species identification and phylogenetic analysis but this approach does not work for all animal species. The New World buckeye butterflies (genus *Junonia*) have several mitochondrial haplotypes that co-occur within species and are shared between species, which makes *cox1* barcodes ineffective for species identification. Previous phylogenetic analysis of *cox1* barcode sequences has suggested that the New World *Junonia* are monophyletic. Complete mitochondrial genome sequences from Old and New World *Junonia* were obtained by next generation sequencing. A total of 18 *Junonia* mitogenomes were used to investigate the phylogeny of the genus and the history of mitochondrial haplotypes in the New World. The widespread *Junonia* haplotype groups A2 and B appear to be derived from two separate invasions of the New World by the Asian-Pacific lineage that gave rise to *J. villida. Junonia* haplotype groups A2 and B became distinct lineages between 1.79and 2.31 million years ago. Both A2 and B haplotypes were present in the New World by approximately 0.96 million years ago. Hybridization and organelle transfer appears to have taken place between the lineages carrying the A2 and B haplotypes. The New World haplotype group A1 found in *J. vestina* is more closely related to the Asian species *J. lemonias* than to other New World haplotypes and represents a third invasion of the New World from the Pacific about 1.60+/-0.32 million years ago. Phylogenetic analysis of DNA barcodes does not accurately reconstruct the evolutionary history of the *Junonia* mitochondrial genome. When *cox1* barcodes do not provide strong phylogenetic resolution due to insufficient numbers of informative sites, whole mitochondrial genome sequencing provides a potentially effective and increasingly inexpensive alternative method for generating phylogenetic data sets.

Keywords: *Junonia*, mitogenomics, phylogenetics of mitochondrial haplotypes, DNA

barcoding, organelle capture

Introduction

In recent years, DNA barcoding, which employs a 658 bp region of *cytochrome oxidase 1 (cox1)*, has become a common method for identifying animal species and for generating datasets for phylogenetic analyses (Hebert et al., 2003, Ratnasingham & Hebert, 2007). Advocates of DNA barcoding have suggested that it is a quick and effective way to identify species based on a standardized DNA sequence, especially in groups where species can be difficult to identify on the basis of morphology (Hebert et al., 2004). Species vary with respect to the degrees of DNA barcode distinctiveness: a barcode sequence divergence of approximately 2% has been used to identify distinct species in birds (Tavares & Baker, 2008). Additionally, bird *cox1* barcodes were correctly identified to species for all 60 sister species pairs in the study (Tavares & Baker, 2008). In freshwater fish, the average amount of genetic distance within the same species was 0.3%, with 87% of the sampled fish being distinguished into species by barcode sequences (Hubert et al., 2008). In the Lepidopteran family Hesperiidae, it has been shown that a difference of only a few nucleotides in the barcode sequence can differentiate between different skipper species (Burns et al., 2007). For some taxa, DNA barcodes are very effective for distinguishing among co-occuring species (Janzen et al., 2005, Hajibabaei et al., 2006). It is less effective in other cases, where only 77% of sampled species were properly identified with barcode sequences (Elias et al., 2007).

Molecular phylogenetics is a standard part of *cox1* barcode analysis, and tools for building trees are integrated into the Barcode of Life Data (BOLD) System (Ratnasingham & Hebert, 2007). Many studies that incorporate DNA barcoding reproduce the trees generated by BOLD to show species relationships (Brévignon & Brévignon, 2012, Pfeiler et al., 2012). To ensure rapid tree construction, BOLD uses a Neighbor-Joining algorithm (Ratnasingham &

Hebert, 2007), which has been criticized because it does not analyze each nucleotide position independently and does not distinguish between synapomorphic and symplesiomorphic character states (Nixon & Wheeler, 1990). Using alternate phylogenetic algorithms (parsimony, maximum likelihood, Bayesian) and methods to determine branch support (bootstrapping, likelihood ratio tests) resolves some of these problems (Brower, 2006). However, because of the relatively short DNA sequence used in DNA barcodes, complete phylogenetic resolution of species using this sequence alone is not always possible (Brower, 2006, Gemmell et al., 2014, Gemmell & Marcus, 2015).

The New World members of the butterfly genus *Junonia* are an informative case study with regards to the limitations of DNA barcoding. The genus *Junonia* is a group of butterflies that occurs in both the Old and New Worlds (Kodandaramaiah & Wahlberg, 2007, Kodandaramaiah, 2009). There are 10 recognized *Junonia* species that occur in the New World (Brévignon, 2004, Brévignon, 2009, Gemmell et al., 2014). Only 4 of the New World species were recognized as full species prior to 2009: *J. coenia, J. evarete, J. genoveva*, and *J. vestina* when Brévignon, 2009 and Brévignon & Brévignon 2012 began elevating subspecies to full species status and describing additional cryptic species from within the *J. evarete* and *J. genoveva* species complexes. Therefore, all historical phylogenetic hypotheses based on phenotype only dealt with these 4 New World taxa. The earliest worker to consider New World *Junonia* phylogeny was Forbes (1928) who hypothesized that *J. coenia, J. evarete*, and *J. genoveva* were all members of a ring species with ongoing gene flow between them, and that the much smaller and morphologically distinct *J. vestina* was the sister taxon to the others. Forbes (1947) later expanded on this hypothesis to suggest that *J. villida* from the Indo-Pacific was the likely sister taxon to all of the New World *Junonia* (Fig. 1-1A). The monophyly of the New

World *Junonia* was uncontested by later workers, but they did suggest alternative Old World sister taxa for the New World members of the genus based on general similarity of color patterns, genital morphology, and behavior: *J. orithya* from Africa (Eliot 1946, Eliot 1947) (Fig 1-1B) or *J. orithya* from Asia (Corbet 1948) (Fig. 1-1C). These authors did not address the relationships within the New World *Junonia*. The first phylogenies of New World *Junonia* using barcode sequences uncovered the existence of several distinct haplotypes among *Junonia* species (Brévignon & Brévignon, 2012, Pfeiler et al., 2012). Additional investigation of New World *Junonia* confirmed that several of these haplotype groups are geographically widespread throughout the Western Hemisphere, and that variation within species can be as much as 3.9% divergence between the *cox1* haplotype sequence variants (Borchers & Marcus, 2014).

The typical pattern in animals is that species consist of individuals, each of which carries a mitochondrial haplotype that belongs to a monophyletic group that is diagnostic for that species (Tavares & Baker, 2008). When disparate clusters of barcodes are found within a single species, this is sometimes a clue that cryptic species are being clustered together within a single taxonomic entity (Burns et al., 2007) or that *Wolbachia* bacterial infection, geographic isolation, or some other mechanism is preventing gene flow between different populations within a species (Kodandaramaiah et al., 2013). When species include individuals carrying distinct mitochondrial haplotype groups, and these haplotype groups are shared between taxa, this is often evidence of either very recent divergence, failure of coalescence, and retention of ancestral mitochondrial variation in emerging taxa (Brower, 1994) or evidence of hybridization between species followed by mitochondrial introgression (Schmidt & Sperling, 2008, Wahlberg et al., 2009). Hybridization is often the preferred hypothesis when the haplotype groups that co-occur within species and are shared between species have divergence that exceeds the 2% threshold often used

to define and identify distinct species (Schmidt & Sperling, 2008, Tavares & Baker, 2008). The true frequency of mitochondrial haplotype sharing between species due to probable hybridization in Lepidoptera is not fully known, but a number of examples have been described in *Grammia* tiger moths (Schmidt & Sperling, 2008), *Papilio* swallowtail butterflies (Dupuis & Sperling, 2015), and *Polygonia* comma butterflies (Wahlberg et al., 2009), suggesting that it may not be a rare phenomenon.

Phylogenetic analysis of mitochondrial sequences in the Old World *Junonia* follows the pattern typical of most animals: species (defined by morphology) are monophyletic (except for *J. orithya*, which may encompass two cryptic species), and mitochondrial haplotypes are not shared between species (Kodandaramaiah & Wahlberg, 2007, Kodandaramaiah, 2009, Gaikwad et al., 2012). Nuclear sequence data for many Old World *Junonia* species is limited, but the data that are available (from (Kodandaramaiah & Wahlberg, 2007, Kodandaramaiah, 2009)) are congruent with the morphology and mitochondrial DNA data. Thus there is no evidence of any retention of ancestral mitochondrial variation or hybridization among Old World species and no reason to expect that mitochondrial gene trees are giving a signal different from the species trees in the Old World *Junonia*. Therefore, mitochondrial gene trees are probably very similar to species trees for the Old World *Junonia*.

The observed pattern is quite different in the New World *Junonia* taxa. Analysis of mitochondrial sequences in the New World *Junonia* shows species (defined by morphology) are not monophyletic (the only New World forms with monophyletic mitochondrial haplotypes are *J. vestina* and *J. coenia grisea*), and mitochondrial haplotypes are shared between species (Brévignon & Brévignon, 2012, Pfeiler et al., 2012, Borchers & Marcus, 2014, Gemmell et al., 2014, Gemmell & Marcus, 2015). Nuclear sequence data support the existence of genetically

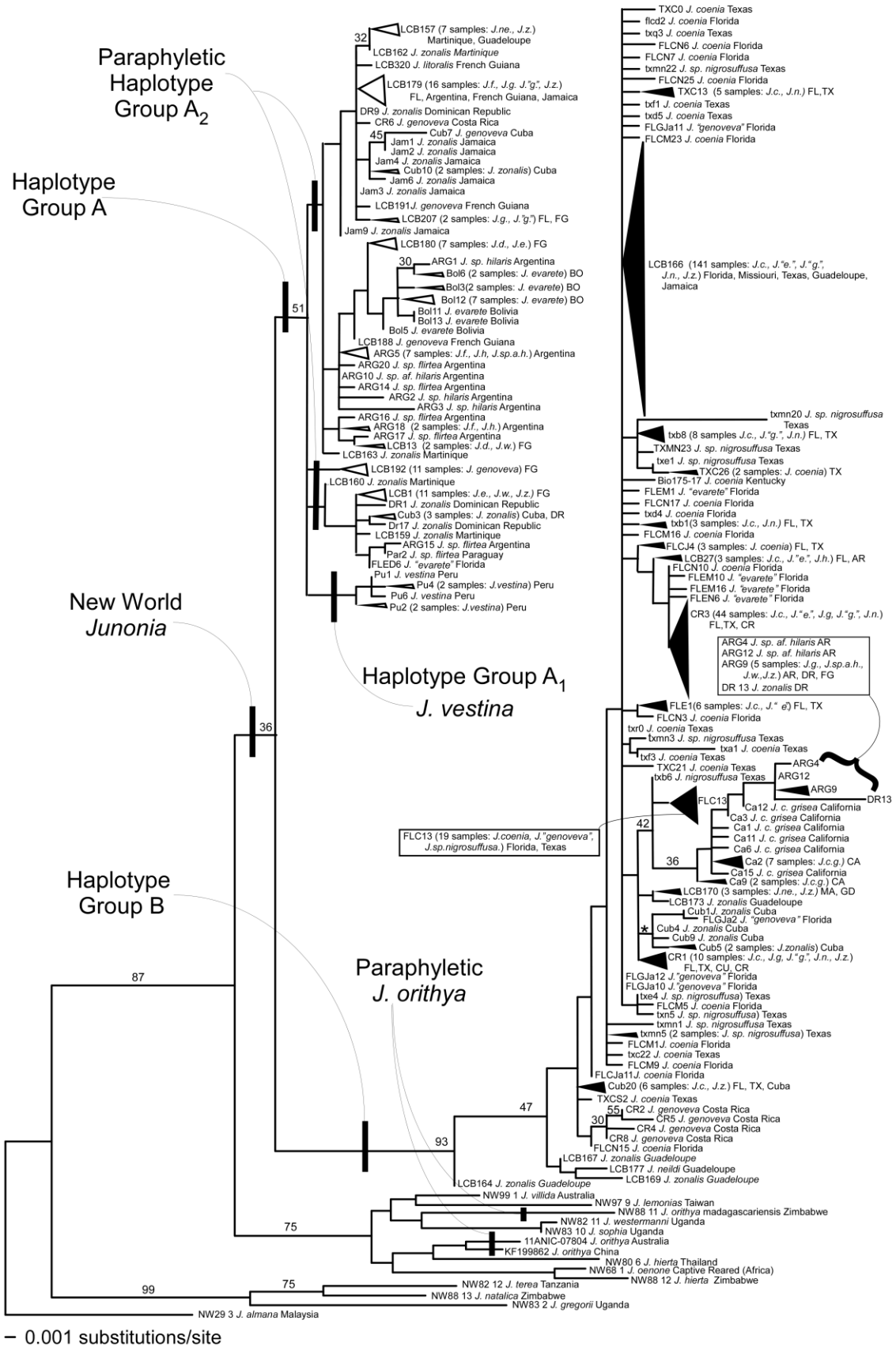
distinct species as identified by morphology and larval host plant use. However, this is not the case for mitochondrial sequence data (again, except for *J. vestina* and *J. coenia grisea*) (Borchers & Marcus, 2014, Gemmell et al., 2014). One interpretation of this pattern is that hybridization between *Junonia* lineages has been widespread in the New World, and so the gene trees based on mitochondrial DNA sequences are not a reliable indicator of the species-level phylogeny for the New World *Junonia* (Gemmell et al., 2014). The possibility of hybridization among New World species is supported by observations of *Junonia* species successfully hybridizing and producing fertile offspring in the laboratory (Hafernik, 1982, Paulsen, 1996), observations of interspecific courtship in the wild (Hafernik, 1982, Minno & Emmel, 1993), and the collection of wild-caught specimens that show color pattern phenotypes that are intermediate between co-occurring *Junonia* species (Hafernik, 1982).

There are four different mitochondrial haplotype groups that occur in New World *Junonia* species. Haplotype group A1 occurs only in *Junonia vestina*, which resides in the Andes Mountains, and has been thought to represent a basal lineage in a monophyletic New World *Junonia* radiation (Pfeiler et al., 2012). Haplotype group A2 is more widespread, occurring throughout South America and the Caribbean, and occurs sporadically in some parts of North America (Pfeiler et al., 2012, Gemmell & Marcus, 2015). Based on the *cox1* barcode region sequences alone, haplotype groups A1 and A2 were named because they were initially thought to be sister clades to each other (haplotype group A), within a larger monophyletic clade that includes all New World *Junonia* species (including haplotype groups A and B) (Pfeiler et al., 2012). This differs from the morphology-based phylogenetic hypotheses about the origins and relationships of the New World *Junonia* species, which place *J. vestina* (and would predict the placement of haplotype group A1) as the sister clade to all other New World *Junonia* (Fig. 1-1A)

(Forbes, 1928, Forbes, 1947).

Haplotype group B predominates in North and Central America, but also occurs in the Caribbean and South America at lower frequency (Brévignon & Brévignon, 2012, Pfeiler et al., 2012, Gemmell & Marcus, 2015). Finally, there is a distinct variant of haplotype group B that occurs only in California in *J. coenia grisea* (Gemmell & Marcus, 2015). However, haplotype groups A2 and B can both be found in many New World *Junonia* species, making it difficult to obtain a phylogeny for *Junonia* species based solely on *cox1* sequences (Gemmell & Marcus, 2015) (Fig. 3-1). The presence of different mitochondrial haplotypes within the same species points to several possibilities: recent speciation, a selective environment that favors variation in mitochondrial genotypes within species, and/or hybridization among *Junonia* species resulting in mitochondrial capture (Borchers & Marcus, 2014). The origins of the New World *Junonia* haplotype groups, their relationships to each other, and their relationships to the Old World *Junonia* haplotypes are somewhat ambiguous based on *cox1* sequences alone, with only weak bootstrap support supporting most of the basal nodes (Gemmell & Marcus, 2015).

Fig. 3-1. Consensus tree of three maximum likelihood trees generated from *cox1* barcode sequences depicting the mitochondrial haplotype groups found in the New World *Junonia* (Gemmell & Marcus, 2015). The single node that differs between these three trees is indicated by an asterisk. Numbers adjacent to nodes indicate bootstrap support. Bootstrap values of less than 30% are omitted. Group A haplotypes occur primarily in South America and the Caribbean. Group A has been further subdivided into A1 and A2 haplotypes. Group B haplotypes occur throughout the Western Hemisphere, but are most common in North and Central America. The *J. orithya* specimens are marked to illustrate that African and Austral-Asia specimens belonging to this species do not form a monophyletic clade. Most *Junonia* species include individuals with both haplotypes. *Junonia vestina* is characterized by the unique haplotype group A1. Only one exemplar of each haplotype was included in the phylogenetic analysis, but the relative abundance of each haplotype is indicated by the vertical width of each triangle. Taxonomic abbreviations: *J.c. (J. coenia), J.d. (J. divaricata), J.e. (J. evarete), J.'e.' (J. 'evarete'), J.f. (J. sp. flirtea), J.g. (J. genoveva), J.'g.' (J. 'genoveva'), J.h. (J. sp. hilaris), J.n. (J. nigrosuffusa), J.ne. (J. neildi), J.sp.a.h. (J. sp. affin hilaris), J.w. (J. wahlbergi) J.z. (J. zonalis)*. Geographic abbreviations: AR (Argentina), BO (Bolivia), CR (Costa Rica), CU (Cuba), FG (French Guiana), FL (Florida, U.S.A.), GD (Guadeloupe), JA (Jamaica), KY (Kentucky, U.S.A.), MA (Martinique), TX (Texas, U.S.A.). Modified from (Gemmell & Marcus, 2015).


Paraphyletic Haplotype Group A2
Haplotype Group A
New World Junonia
Haplotype Group A1
J. vestina
Haplotype Group B
Paraphyletic J. orithya
LCB157 (7 samples: J.ne., J.z.) Martinique, Guadeloupe
LCB162 J. zonalis Martinique
LCB320 J. litoralis French Guiana
LCB179 (16 samples: J.f., J.g. J."g", J.z.) FL, Argentina, French Guiana, Jamaica
DR9 J. zonalis Dominican Republic
CR6 J. genoveva Costa Rica
Cub7 J. genoveva Cuba
Jam1 J. zonalis Jamaica
Jam2 J. zonalis Jamaica
Jam4 J. zonalis Jamaica
Cub10 (2 samples: J. zonalis) Cuba
Jam6 J. zonalis Jamaica
Jam3 J. zonalis Jamaica
LCB191J. genoveva French Guiana
LCB207 (2 samples: J.g., J."g") FL, FG
Jam9 J. zonalis Jamaica
LCB180 (7 samples: J.d., J.e.) FG
ARG1 J. sp. hilaris Argentina
Bol6 (2 samples: J. evarete) BO
Bol3(2 samples: J. evarete) BO
Bol12 (7 samples: J. evarete) BO
Bol11 J. evarete Bolivia
Bol13 J. evarete Bolivia
Bol5 J. evarete Bolivia
LCB188 J. genoveva French Guiana
ARG5 (7 samples: J.f., J.h, J.sp.a.h.) Argentina
ARG20 J. sp. flirtea Argentina
ARG10 J. sp. af. hilaris Argentina
ARG14 J. sp. flirtea Argentina
ARG2 J. sp. hilaris Argentina
ARG3 J. sp. hilaris Argentina
ARG16 J. sp. flirtea Argentina
ARG18 (2 samples: J.f., J.h.) Argentina
ARG17 J. sp. flirtea Argentina
LCB13 (2 samples: J.d., J.w.) FG
LCB163 J. zonalis Martinique
LCB192 (11 samples: J. genoveva) FG
LCB160 J. zonalis Martinique
LCB1 (11 samples: J.e., J.w., J.z.) FG
DR1 J. zonalis Dominican Republic
Cub3 (3 samples: J. zonalis) Cuba, DR
Dr17 J. zonalis Dominican Republic
LCB159 J. zonalis Martinique
ARG15 J. sp. flirtea Argentina
Par2 J. sp. flirtea Paraguay
FLED6 J. "evarete" Florida
Pu1 J. vestina Peru
Pu4 (2 samples: J.vestina) Peru
Pu6 J. vestina Peru
Pu2 (2 samples: J.vestina) Peru
TXC0 J. coenia Texas
flcd2 J. coenia Florida
txq3 J. coenia Texas
FLCN6 J. coenia Florida
FLCN7 J. coenia Florida
txmn22 J. sp. nigrosuffusa Texas
FLCN25 J. coenia Florida
TXC13 (5 samples: J.c., J.n.) FL,TX
txf1 J. coenia Texas
txd5 J. coenia Texas
FLGJa11 J. "genoveva" Florida
FLCM23 J. coenia Florida
LCB166 (141 samples: J.c., J."e.", J."g.", J.n., J.z.) Florida, Missiouri, Texas, Guadeloupe, Jamaica
txmn20 J. sp. nigrosuffusa Texas
txb8 (8 samples J.c., J."g.", J.n.) FL, TX
TXMN23 J. sp. nigrosuffusa Texas
txe1 J. sp. nigrosuffusa Texas
TXC26 (2 samples: J. coenia) TX
Bio175-17 J. coenia Kentucky
FLEM1 J. "evarete" Florida
FLCN17 J. coenia Florida
txd4 J. coenia Florida
txb1(3 samples: J.c., J.n.) FL, TX
FLCM16 J. coenia Florida
FLCJ4 (3 samples: J. coenia) FL, TX
LCB27(3 samples: J.c., J."e.", J.h.) FL, AR
FLCN10 J. coenia Florida
FLEM10 J. "evarete" Florida
FLEM16 J. "evarete" Florida
FLEN6 J. "evarete" Florida
CR3 (44 samples: J.c., J."e.", J.g, J."g.", J.n.) FL,TX, CR
ARG4 J. sp. af. hilaris AR
ARG12 J. sp. af. hilaris AR
ARG9 (5 samples: J.g., J.sp.a.h., J.w.,J.z.) AR, DR, FG
DR 13 J. zonalis DR
FLE1(6 samples: J.c., J." e") FL, TX
FLCN3 J. coenia Florida
txr0 J. coenia Texas
txmn3 J. sp. nigrosuffusa Texas
txa1 J. coenia Texas
txf3 J. coenia Texas
TXC21 J. coenia Texas
txb6 J. nigrosuffusa Texas
ARG4
ARG12
ARG9
DR13
FLC13
Ca12 J. c. grisea California
Ca3 J. c. grisea California
Ca1 J. c. grisea California
Ca11 J. c. grisea California
Ca6 J. c. grisea California
Ca2 (7 samples: J.c.g.) CA
Ca15 J. c. grisea California
Ca9 (2 samples: J.c.g.) CA
FLC13 (19 samples: J.coenia, J."genoveva", J.sp.nigrosuffusa.) Florida, Texas
LCB170 (3 samples: J.ne., J.z.) MA, GD
LCB173 J. zonalis Guadeloupe
Cub1 J. zonalis Cuba
FLGJa2 J. "genoveva" Florida
Cub4 J. zonalis Cuba
Cub9 J. zonalis Cuba
Cub5 (2 samples: J.zonalis) Cuba
CR1 (10 samples: J.c., J.g, J."g.", J.n., J.z.) FL,TX, CU, CR
FLGJa12 J."genoveva" Florida
FLGJa10 J."genoveva" Florida
txe4 J. sp. nigrosuffusa) Texas
FLCM5 J. coenia Florida
txn5 J. sp. nigrosuffusa) Texas
txmn1 J. sp. nigrosuffusa Texas
txmn5 (2 samples: J. sp. nigrosuffusa) Texas
FLCM1J. coenia Florida
txc22 J. coenia Texas
FLCM9 J. coenia Florida
FLCJa11J. coenia Florida
Cub20 (6 samples: J.c., J.z.) FL, TX, Cuba
TXCS2 J. coenia Texas
CR2 J. genoveva Costa Rica
CR5 J. genoveva Costa Rica
CR4 J. genoveva Costa Rica
CR8 J. genoveva Costa Rica
FLCN15 J. coenia Florida
LCB167 J. zonalis Guadeloupe
LCB177 J. neildi Guadeloupe
LCB169 J. zonalis Guadeloupe
LCB164 J. zonalis Guadeloupe
NW99 1 J. villida Australia
NW97 9 J. lemonias Taiwan
NW88 11 J. orithya madagascariensis Zimbabwe
NW82 11 J. westermanni Uganda
NW83 10 J. sophia Uganda
11ANIC-07804 J. orithya Australia
KF199862 J. orithya China
NW80 6 J. hierta Thailand
NW68 1 J. oenone Captive Reared (Africa)
NW88 12 J. hierta Zimbabwe
NW82 12 J. terea Tanzania
NW88 13 J. natalica Zimbabwe
NW83 2 J. gregorii Uganda
NW29 3 J. almana Malaysia
32
45
51
30
36
42
36
47
55
30
93
87
75
99
75
0.001 substitutions/site

By expanding the available data set by obtaining full-length mitochondrial genome sequences from New and Old World *Junonia*, I aim to investigate the origins and relationships of the New World haplotypes. The genus *Junonia* originated and diversified in Africa, and then spread to a secondary center of diversification in South Asia (Kodandaramaiah & Wahlberg, 2007). There are two hypotheses for how *Junonia* came to the New World: the single invasion event hypothesis and the multiple invasion event hypothesis (Gemmell & Marcus, 2015). In the single invasion event hypothesis, New World *Junonia* are monophyletic and all speciation and haplotype diversification has occurred in the New World after the arrival of the colonizing *Junonia* population. If this is the case, the prediction would be that the A and B mitochondrial haplotype groups are sister taxa and monophyletic with respect to the Old World mitochondrial genotypes. This has been the working hypothesis for most *Junonia* taxonomists (Figs. 1-1, 3-1) (Forbes, 1928, Eliot, 1946, Eliot, 1947, Forbes, 1947, Eliot, 1949, Kodandaramaiah & Wahlberg, 2007, Neild, 2008, Kodandaramaiah, 2009, Pfeiler et al., 2012).

In the multiple invasion hypothesis, the A and B haplotypes groups are the result of two separate *Junonia* invasions to the New World (Gemmell & Marcus, 2015). While their phylogenetic analyses produced a monophyletic New World *Junonia* similar to previous workers (Fig. 1-1D) (Kodandaramaiah & Wahlberg, 2007, Kodandaramaiah, 2009, Pfeiler et al., 2012), the bootstrap support for the basal node of the clade was very weak (36%) and haplotype network analysis suggested a stronger association between *J. villida* and haplotype group A, than the association between haplotype groups A and B (Gemmell & Marcus, 2015). Gemmell and Marcus (2015) therefore proposed the two invasion hypothesis as an alternative to be considered. In this scenario, the New World haplotypes will not be monophyletic and some of the haplotypes will be more closely related to haplotypes found in Old World *Junonia* species (which with few

exceptions (*J. orithya*)) are species specific, haplotypes within a species are monophyletic, and thus Old World haplotypes can be referred to unambiguously by the species name (Fig. 3-1)) than they are to other New World Haplotypes. In the scenario proposed by Gemmell and Marcus (2015) based on haplotype network analysis, as an alternative to New World monophyly, haplotype groups A1 and A2 are sister clades derived from one invasion, and haplotype group B (including the California variant) are derived from a second invasion. The order in which these invasions might have occurred was unclear based on the *cox1* barcode data set (Gemmell & Marcus, 2015)

Additionally, regardless of whether there have been one or multiple invasion events, it has yet to be determined whether *Junonia* colonized the New World by a trans-Pacific route from Austral-Asia or by a trans-Atlantic route from Africa (Kodandaramaiah & Wahlberg, 2007). A study investigating the origin of *Junonia* by examining *cox1, ef1-alpha,* and *wingless* coding sequences was unable to determine the Old World source for New World colonists (Kodandaramaiah & Wahlberg, 2007). Two different sister species for the New World *Junonia* radiation have been proposed: *J. orithya* and *J. villida* (Eliot, 1946, Eliot, 1947, Forbes, 1947, Corbet, 1948, Eliot, 1949). *Junonia orithya* is found in Africa and Asia, while *J. villida* has an Asian-Pacific distribution (Vane-Wright & Tennent, 2011). Thus, the first *Junonia* colonists of the New World may have come from across either the Atlantic or the Pacific Oceans. A haplotype network analysis suggested that *J. villida* may be the sister taxon to the New World *Junonia* on the basis of *cox1* haplotypes, but this analysis only included *J. orithya* from China and Uganda and *J. villida* from Australia as possible Old World outgroups and did not include additional Old World species (Gemmell & Marcus, 2015).

Due to the likely history of hybridization and mitochondrial introgression among New

World *Junonia* taxa, phylogenetic analysis of complete mitochondrial genome sequences may not produce a definitive species-level phylogeny for the New World members of this genus. However, analyses of these data will provide information about the number of invasion events, the geographic origin(s) of the source population(s), the timing of invasion and diversification events, and a test of the monophyly of the New World *Junonia*. As such, it will provide important new information about the evolutionary genetic history of *Junonia* in the New World.

Finally, the New World *Junonia* is a useful test case for the use of *cox1* barcodes for species identification and phylogenetics, which are the two most common applications of DNA barcoding techniques (Ratnasingham & Hebert, 2007). Previous studies that compared *cox1* barcodes with species as defined by morphology (Brévignon & Brévignon, 2012, Pfeiler et al., 2012, Borchers & Marcus, 2014, Gemmell et al., 2014, Gemmell & Marcus, 2015), larval host plant use (Brévignon & Brévignon, 2012, Gemmell et al., 2014), or nuclear DNA markers (Borchers & Marcus, 2014, Gemmell & Marcus, 2015) in the New World *Junonia* already suggest that mitochondrial barcodes cannot be used for reliable species identification, and therefore would not be appropriate for species-level phylogenetic studies. Thus, *cox1* based gene trees would not be expected to match New World *Junonia* species trees. A question that remains is whether DNA barcode sequences contain sufficient phylogenetic information to accurately reflect the evolutionary history of the mitochondrial genomes of which they are a sample (Funk & Omland, 2003, Schmidt & Sperling, 2008, Spooner, 2009). If it is shown that phylogenetic trees based on *cox1* barcodes do not accurately reflect the phylogeny of the mitochondrial genome (let alone the phylogeny of the organisms carrying the mitochondrial genome), the value of the now common practice of using *cox1* barcodes alone to estimate phylogenetic relationships among most taxa will be brought into question.

Objectives:

To use complete Old and New World *Junonia* mitochondrial genome sequences to achieve the following objectives:

1. To create a robust phylogeny of the mitochondrial haplotypes found in the New World members of genus *Junonia* in order to examine relationships between the different mitochondrial haplotypes.

Hypothesis 1A: The New World *Junonia* mitochondrial haplotypes are monophyletic.

Prediction: All New World *Junonia* haplotypes form a single clade that does not include any Old World haplotypes.

Hypothesis 1B: The New World *Junonia* are not monophyletic.

Prediction: The New World *Junonia* haplotypes do not form a single clade and are interspersed with Old World haplotypes.

2. To determine the closest Old World relative(s) to the New World *Junonia*.

Hypothesis 2A: *Junonia villida* from the Asia-Pacific region is the sister taxon to the New World *Junonia* as proposed by Forbes (1947.)

Prediction: *Junonia villida* is the sister clade to haplotypes from the New World *Junonia*.

Hypothesis 2B: *Junonia orithya madagascariensis* from Africa is the sister taxon to the New World *Junonia* as proposed by Eliot (1946, 1947).

Prediction: *Junonia orithya madagascariensis* is the sister clade to haplotypes from the New World *Junonia.*

Hypothesis 2C: *Junonia orithya orithya* from Asia is the sister taxon to the New World *Junonia* as proposed by Corbet (1948).

Prediction: *Junonia orithya orithya* is the sister clade to haplotypes from the New World *Junonia.*

Hypothesis 2D: A taxon other than those listed above is the sister taxon to the New World *Junonia*.

Prediction: Neither *J.villida*, *J. orithya madagascariensis*, nor *J. orithya orithya* are the sister clade to haplotypes from the New World *Junonia.*

3. To determine how many times *Junonia* invaded the New World.

Hypothesis 3A: *Junonia* invaded the New World once.

Prediction: The New World *Junonia* mitochondrial haplotypes are monophyletic and share a common sister taxon.

Hypothesis 3B: *Junonia* invaded the New World twice as proposed by Gemmell and Marcus (2015).

Prediction: The New World *Junonia* mitochondrial haplotypes form two different clades that may have different Old World sister taxa.

Hypothesis 3C: *Junonia* invaded the New World three or more times.

Prediction: The New World *Junonia* mitochondrial haplotypes form three or more different clades that may have different Old World sister taxa.

4. To use the amount of divergence between the New World *Junonia* mitochondrial haplotype groups to establish time frames for the origin of each groups.

This objective is exploratory and no specific hypotheses are being tested.

Methods

Sample collection and DNA extraction

Junonia specimens were collected from the wild by handheld butterfly net (Table 3-1). Specimens were killed by pinching, and placed in glassine envelopes at room temperature for shipping. When the specimens arrived in the laboratory, their identities were confirmed by identification of morphological characters (Forbes, 1928, Corbet, 1948, Munroe, 1951, Tilden, 1970, Turner & Parnell, 1985, Larsen, 1991, Igarishi & Fukada, 1997, Austin & Emmel, 1998, Igarishi & Fukada, 2000, Neild, 2008, Kodandaramaiah, 2009, Vane-Wright & Tennent, 2011, Brévignon & Brévignon, 2012, Mecenero et al., 2013, Borchers & Marcus, 2014). The specimens were then were stored at -20ºC. Specimens included in this analysis were selected in order to test each of the phylogenetic hypotheses that have been proposed for the origins of the New World *Junonia* (Eliot, 1946, Eliot, 1947, Forbes, 1947, Corbet, 1948, Eliot, 1949), and to explore the genetic diversity within the major New World haplotype groups A and B. Six of the currently recognized New World species were included in this analysis, along with two *Junonia* forms of indeterminate taxonomic status. The remaining four species share typical (and near identical) A2 and B mitochondrial haplotypes with other New World species that were sequenced, so their absence is very unlikely to have a major impact on phylogenetic reconstruction. DNA was prepared from one leg of each specimen using the DNEasy Blood and Tissue kit and a Qiacube extraction robot (Qiagen) with modifications to the standard animal tissue extraction protocol for total DNA as described previously (Gemmell & Marcus, 2015) (Chapter 2). DNA yield and quality was checked on a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, Delaware, USA) and a Qubit 2.0 fluorometer (Life Technologies, Carlsbad, California, USA). Extracted DNA was stored at -20°C. Specimens are

currently housed in the research collection of the Marcus laboratory at the University of Manitoba so that they can be readily consulted for ongoing follow-up studies by other members of the laboratory, but will be provided as vouchers to a public entomology collection when these studies are complete.

Table 3-1. Old World and New World mitochondrial genomes sequenced in this study.

Species	**Voucher**	**Origin, Collection Date, Collector**	**Mitochondrial Genome length**	**Citation/ Haplotype**
Old World				
J. almana	CNA10	China: Guangxi, Guilin July-2012 collector unknown (purchased specimen)	15,196	
J. lemonias	CNL3	China: Yunnan, Yuanjing 20-May-2014 Robert Dowell	15,230	Chapter 2, McCullagh & Marcus, 2015
J. orithya madagascariensis	MWO12	Malawi: Mt. Mulanje, Fort Lister 03-Jun-2007 R. J. Murphy	15,206	
J. villida	Aus1	Australia: New South Wales, Sydney 15-Dec-1999 collector unknown (purchased specimen)	15,194	
New World				
J. coenia	KY24	USA: Kentucky, Hart County 01-Oct-2007 Jeffrey M. Marcus	15,210	B
J. coenia	TXMC8	USA: Texas, Cameron County 28-May-2008 Jeffrey M. Marcus	15,209	B

J. coenia grisea	CA7	USA: California, Sacramento County 26-Aug-2010 Robert Dowell	15,213	B (California variant)
J. genoveva	FLGJa1	USA: Florida, Dade County 11-Jan-2007 Jeffrey M. Marcus	15,214	A2
J. genoveva	FGG2015.4	French Guiana: Saül 07-Nov-2014 Christian Brévignon	14,858 (partial)	A2
J. "genoveva" (mangrove feeder)	EP26	Mexico: Sonora, Estero del Soldado 26-Feb-2015 Edward Pfeiler	15,212	B
J. litoralis	FGL2015.1	French Guiana: Macouria 31-Jan-2015 Christian Brévignon	15,213	A2
J. sp. flirtea (light phenotype)	Arg18	Argentina: Colegialis Jardin, Buenos Aires 29-Jan-2010 Rudi Mattoni	15,225	A2
J. sp. flirtea (dark phenotype)	Arg20	Argentina: Buenos Aires 28-Feb-2010 Rudi Mattoni	15,210	A2
J. sp. hilaris	Arg4	Argentina: Buenos Aires 28-Feb-2010 Rudi Mattoni	15,204	B
J. vestina	PU6	Peru: Cuzco, Wayquecha Research Station 30-Oct-2010 Loran Gibson	15,224	A1
J. zonalis	Cubz2015.1	Cuba: Pinar del Río 03-Jan-2015 Carlos A. Cruz	14,118 (partial)	A2

DNA preparation and library construction

The extracted *Junonia* DNA samples then underwent additional processing at the Next Generation Sequencing Platform facility at the Manitoba Institute of Child Health. The DNA sample was sheared by sonication with an S220 Focused-Ultrasonicator (Covaris, Woburn, Massachusetts, USA) to an average fragment size of 500 bp. Fragment sizes were evaluated using a High Sensitivity DNA chip for the Bioanalyzer 2100 electrophoresis system (Agilent, Santa Clara, California, USA) using the standard manufacturer protocol. A KAPA LTP Library Preparation kit (Kapa Biosystems, Boston, Massachusetts, USA) was used to prepare an indexed library from the sheared samples for loading onto a MiSeq NextGen Sequencing Instrument (Illumina, San Diego, California, USA). Two experiments with 10 and 12 indexed libraries (respectively) were processed simultaneously on this instrument. The MiSeq reagent V3 300X2 paired end reagent kit (Illumina) was used to perform the sequencing.

Genome assembly and phylogenetic analysis

Sequence data for each index were aligned to a previously published and annotated *J. orithya* complete mitochondrial genome reference sequence (Shi et al., 2013) using Geneious 8.1.3 software (Kearse et al., 2012). After three assembly iterations were performed, no additional fragments were added to the contig, so the software automatically terminated the assembly. Once assembly was complete, the annotations from a *J. lemonias* reference sequence were transferred with minor adjustments (Chapter 2) onto the *Junonia* mitochondrial genome sequence from each index using Geneious.

Additional mitochondrial genomes from nymphalid butterfly species were identified for comparison purposes on the basis of available full-length mitochondrial genomes in GenBank (Table 3-2). Sequences from the entire mitochondrial genome were aligned in the Clustal Omega multiple sequence alignment program (Sievers et al., 2011) using the default settings and converted to NEXUS file format. (Darriba et al., 2012) As in a previous study, while there are several indels in the intergenic regions of the aligned data set, (Chapter 2, (McCullagh & Marcus, 2015)) they were of consistent length and sequence position among all samples, so there were no regions of ambiguous alignment in the entire data set. The phylogenetic analysis was conducted in PAUP* 4.0b8/4.0d78 (Swofford, 2002). Unless otherwise specified, default settings were used in PAUP. For parsimony-based phylogenetic analysis, 1 million heuristic search replicates were performed with random sequence and tree bisection and reconnection branch swapping on best trees only, multiple trees saved at each step, and retaining all best trees. This was followed by a bootstrap analysis with 1 million random sequence addition fast addition search replicates, retaining all groups compatible with 50% bootstrap consensus. (Swofford, 2002)

To test the amount of useful data contained in barcode sequences, I also performed a phylogenetic analysis using barcode sequences from the published genomes (Table 3-2) and the sequenced mitochondrial genomes (Table 3-3). Parsimony analysis of barcode *cox1* sequences was conducted with PAUP* using 1 million random sequence addition search replicates. PAUP* settings were default with multiple trees saved at each step and a swap performed on best trees only.

The degree of bootstrap support for phylogenetic clades was classified as poor (<55%), weak (55-64%), moderate (65-74%), good (75-84%), very good (85-94%), or strong (95-100%)

(Hillis & Bull, 1993, Huelsenbeck et al., 1996). Based on simulation studies, 88% or more bootstrap support corresponds to a 95% confidence limit that the data support a particular node (Zander, 2004).

Table 3-2. Published mitochondrial genomes used in phylogenetic analyses.

Species	**GenBank Accession number**	**Origin**	**Reference**
Hypolimnas bolina	NC_026072	China	Shi et al., 2015
Junonia almana	NC_024407	Japan	Wu et al., 2014
Junonia lemonias	KP941756	China	Chapter 2, McCullagh & Marcus, 2015
Junonia orithya	NC_022697	China	Shi et al., 2013
Yoma sabina	NC_024403	SE Asia	Wu et al., 2014

Table 3-3. Next generation sequencing statistics for sequenced mitochondrial genomes.

Species	**Voucher**	**# of reads**	**Aligned reads**	**Mean coverage**	**Minimum coverage**	**Maximum coverage**
J. almana	CNA10	8,248,672	19,994	276.9	35	301
J. orithya madagascariensis	MWO12	4,488,306	2,302	31.1	2	102
J. villida	Aus1	4,612,172	312	5.0	1	24
J. vestina	PU6	3,444,228	7,832	73.5	31	176
J. coenia	KY24	8,861,870	1,018	14.3	2	35
J. coenia	TXMC8	4,359,310	3,348	48.0	1	148
J. coenia grisea	CA7	7,025,478	16,691	249.7	1	462
J. genoveva	FLGJa1	7,711,580	9,487	137.5	65	209
J. genoveva	FGG2015.4	4,753,892	5,874	82.7	1	233
J. "genoveva" (mangrove feeder)	EP26	4,945,286	6,234	87.9	2	345
J. litoralis	FGL2015.1	5,789,224	4,644	63.7	1	233
J. sp. flirtea (light phenotype)	Arg18	6,235,752	22,783	314.0	69	498
J. sp. flirtea (dark phenotype)	Arg20	5,627,702	12,423	150.8	1	372
J. sp. hilaris	Arg4	7,994,138	7,898	117.7	26	214
J. zonalis	Cubz2015.1	3,428,852	2,243	33.7	1	92

The best likelihood models for analysis of full length mitochondrial genomes and for the 658 bp *cox1* barcode sequence (Hebert et al., 2003) were determined to be a General Time Reversible (GTR) model with models for the proportion of invariable sites (I) and for the gamma distribution (G) using jModeltest 2.1.7 (Darriba et al., 2012) and a likelihood ratio test (Huelsenbeck & Rannala, 1997). For the maximum likelihood (GTR+I+G model) algorithm, the

parameters I and G were estimated in PAUP for both the barcode sequences and the complete mitochondrial genomes for maximum likelihood phylogenetic analyses with GTR model settings using 10 heuristic search replicates with random sequence and tree bisection and reconnection branch swapping on best trees only, and retaining all best trees (otherwise all PAUP default settings were used). For the whole mitochondrial genome, the proportion of invariable sites (I) was found to be 0.623741 and the estimated gamma (G) was 0.753208. For the *cox1* barcode sequence, the proportion of invariable sites (I) was 0.259968 and the estimated gamma (G) was 0.155801. The barcode sequences were then reanalyzed (10 maximum likelihood heuristic search replicates as above) using the I and G values calculated for the whole mitochondrial genome and an identical tree topology was recovered using both estimates for I and G. For this reason, the whole mitochondrial genome estimates for I and G were used for subsequent maximum likelihood analyses. For the barcode sequences, the complete mitochondrial genomes, and for the complete mitochondrial genomes minus the barcode sequences, 1 million maximum likelihood heuristic search replicates with random sequence addition and tree bisection and reconnection branch swapping on best trees only, multiple trees saved at each step, ad retaining all best trees. This was followed by a bootstrap analysis with 1 million random sequence addition fast addition search replicates, and retaining all groups compatible with 50% bootstrap consensus.

Calculating divergence times

Divergence times between selected haplotypes was calculated using a molecular clock and the EMBOSS Needle to calculate % sequence divergence between species pairs (Rice et al., 2000). First, the number of identical positions was determined and converted into a percentage. Then, the percent divergence was calculated by subtracting the percent similarity from 100. Estimates and confidence intervals for the age of divergence for *Junonia* species pairs were

obtained by using a linear regression of date calibrated arthropod divergence data from Brower (1994). It should be noted that while Brower (1994) calibrated the molecular clock using divergence data for only portions of the mitochondrial genome (with an emphasis on *cox1*) and my data set consists of complete mitochondrial genomes, percent divergence between A2 and B haplotype groups for the *cox1* barcode fragment (Borchers & Marcus, 2014) and A2 and B haplotype groups for the entire mitochondrial genome (Table 3-3) are nearly identical in *Junonia*. This suggests that the Brower (1994) molecular clock calibration regression is a reasonable molecular clock function for use with the *Junonia* mitochondrial genome data set, particularly since there are no available nodes within the data set with firm dates that can be used for internal calibration.

Results

Ten complete New World and four complete Old World *Junonia* mitochondrial genomes were sequenced (Table 3-3). Two partial (greater than 90% sequenced) New World mitogenomes were also recovered from our experiments (Table 3-3). One of the Old World *Junonia* mitochondrial genomes (*J. lemonias*) from this experiment has been published separately (Chapter 2, McCullagh & Marcus, 2015). Prior to this work, only two other complete Old World *Junonia* mitochondrial genomes and no New World mitochondrial genomes were available (Shi et al., 2013, Wu et al., 2014). The remaining six *Junonia* samples included in this experiment did not produce complete or nearly complete mitochondrial genome sequences (potentially due to poor DNA quality) and so were set aside and not included in further analyses. Thus a total of 18 *Junonia* mitochondrial genome sequences, plus one sequence each from the genera *Yoma* and *Hypolimnas* were used as outgroups for phylogenetic analysis. The total length

of the aligned data set was 15,477 bp with 1,311 variable charcters and 1,606 parismony informative characters.

Only one parismony tree using barcode *cox1* sequences was recovered with a score of 193 steps (Table 3-4). 926,476/1 million replicates recovered the identical parsimony barcode tree.

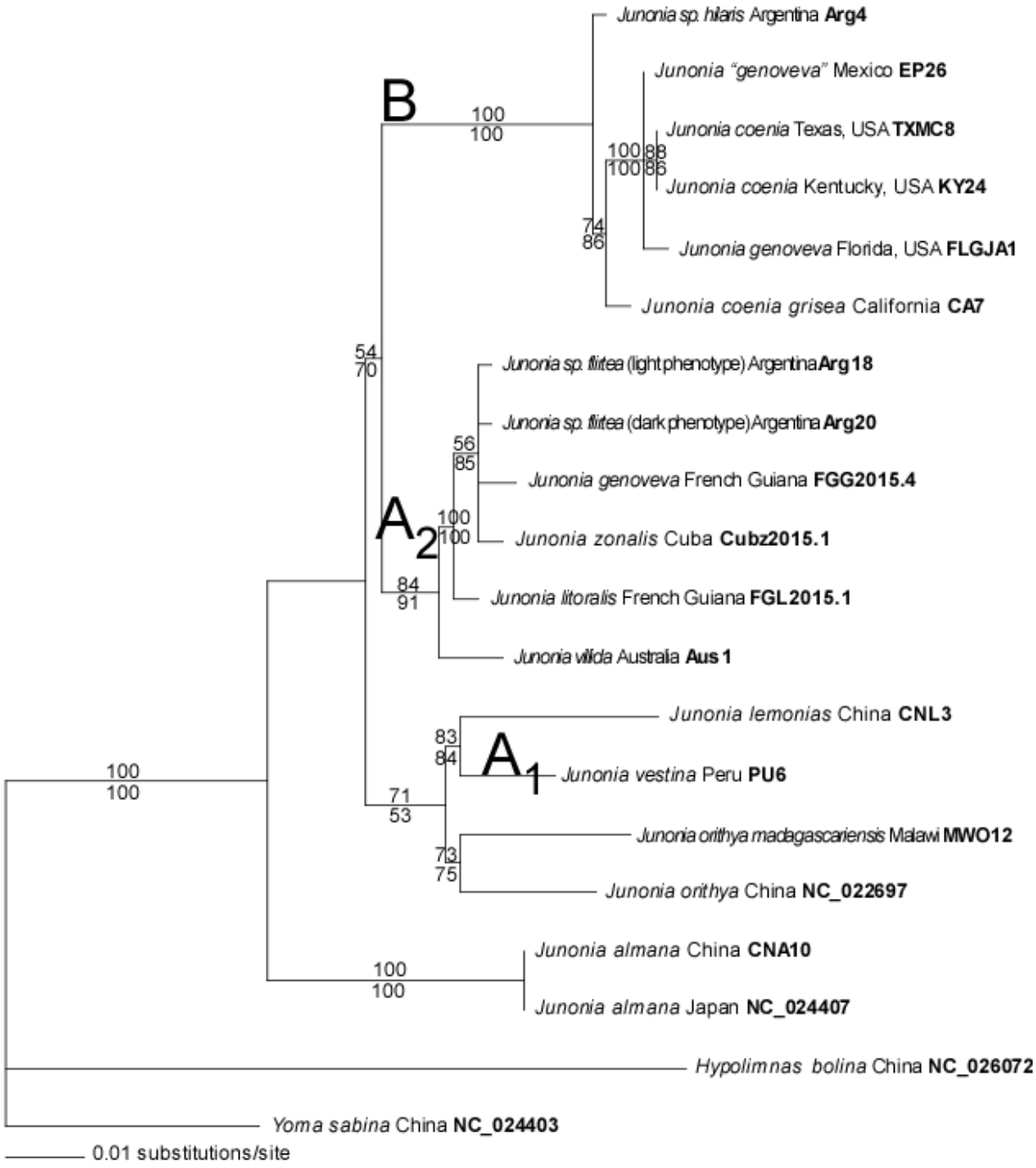
Table 3-4. Parsimony tree islands results found in 1 million random addition heuristic search replicates of the whole mitochondrial genome data set. Note that under the setting used for my analyses, PAUP does not retain suboptimal trees between heuristic search replicates and for this reason cannot accurately calculate the number of suboptimal trees found in each island. It does retain all trees of the optimal score. In this case, the same tree of length 193 was found in 92.6% of 1 million search replicates.

Island	**Size**	**First tree**	**Last tree**	**Score**	**First replicate**	**Times hit**
1	1	1	1	193	1	926476
2	1	-	-	194	726	2064
3	0	-	-	194	10	68195
4	1	-	-	195	196594	1
5	0	-	-	195	3011	1378
6	0	-	-	196	1157	1866
7	0	-	-	197	53487	14
8	0	-	-	198	146298	6

One optimal phylogenetic tree was produced by both the parsimony and maximum likelihood search algorithms of the full mitochondrial genome data set, each with the same topology (Fig. 3-2). The best maximum likelihood tree had a likelihood score of 46103.192 and the shortest parsimony tree had a score of 4961 steps. Maximum likelihood and parsimony bootstrap support for almost all of the nodes was very robust (Fig. 3-2). All of the B haplotypes grouped together with strong bootstrap support. Similarly, the species in haplotype group A2 also grouped together with strong bootstrap support. *Junonia villida*, from Australia, is the sister clade to haplotype group A2 with very good parsimony and good maximum likelihood bootstrap support. Haplotype groups A2 (including *J. villida*) and B form a clade with 54/70 maximum likelihood/parsimony bootstrap support. However, haplotype A1, *J. vestina* from Peru, is the sister clade to *J. lemonias* from China, again with good bootstrap support (83% maximum likelihood, 84% parsimony). The outgroup to this species pair are *J. orithya* and *J.orithya madagascariensis.* This outgroup relationship of *J. orithya* and *J. orithya madagascariensis* has 73% maximum likelihood and 75% parsimony bootstrap support. . Divergence times between clades of interest were calculated and are reported in Table 3-5. From the calculations of divergence times, it can be inferred that the lineages that gave rise to haplotype group B and to *J. villida* and haplotype group A2 diverged less than 2.75 million years ago. Haplotype group B appears to have started diversifying within the New World within the last 1.5 million years based on the most divergent haplotypes sampled. Haplotype group A2 split from its Indo-Pacific sister taxon *J. villida* less than 1.9 million years ago and based on the haplotypes sampled, A2 has been diversifying within the New World for the last 1.3 million years. New World *J. vestina* also appears to have separated from its Asian sister taxon *J. lemonias* less than 1.9 million years ago.

Similarly, the Asian and African subspecies of *J. orithya* diverged from each other within the last 2 million years.

Fig. 3-2. Maximum likelihood tree of complete mitochondrial genomes. Two partial mitochondrial genomes (Cubz2015.1 and FGg2015.4) were also included. Parsimony analysis produced a single tree with an identical tree topology. Maximum likelihood bootstrap values are above the nodes, and parsimony bootstrap values are below the nodes.



In addition to the analysis of the complete mitochondrial genomes, I also analyzed these data in smaller subsets. Parsimony bootstrap analysis was conducted in five subsets: *cox1*

barcode, *cox1* coding sequence, mitochondrial genomes with *cox1* removed, mitochondrial genomes without *cox1* barcode sequence, and whole mitochondrial genomes (Fig. 3-3). The *cox1* barcode fragment parsimony bootstrap tree is unable to resolve the relationships among many New World haplotypes: haplotype group B is monophyletic, but most of the component haplotypes are part of a large unresolved polytomy, and New World haplotype A1 is embedded in a large polytomy containing both Old World species and New World haplotypes (Fig. 3-3 A). As more data is added to the analysis, both of these polytomies become increasingly well resolved, and the bootstrap support for nodes increases steadily (Fig. 3-3 B, C, D, E). A direct comparison of panel A and panel E shows that the whole mitochondrial genomes (15477 bp) provides a much better resolved and supported phylogenetic hypothesis than *cox1* barcode fragments (658 bp). An essentially similar pattern was produced by maximujm likelihood analysis of the barcode fragment, the mitochondrial genomes except for the barcode fragment, and the whole mitochondrial genomes (Fig. 3-4).

Fig. 3-3. Parsimony bootstrap analysis of genome data divided into subsets. Panel A: parsimony bootstrap of *cox1* barcode fragment. Panel B: parsimony bootstrap of *cox1* coding sequence. Panel C: parsimony bootstrap of mitochondrial genomes excluding *cox1*. Panel D: parsimony bootstrap of mitochondrial genomes excluding *cox1* barcode fragment. Panel E: parsimony bootstrap of whole mitochondrial genomes. Specimens denoted with an * are reference sequences downloaded from GenBank. *J. orithya*: accession number NC_022697. *J. almana*: accession number NC_024407. *H. bolina*: accession number NC_026072. *Y. sabina*: accession number NC_024403.

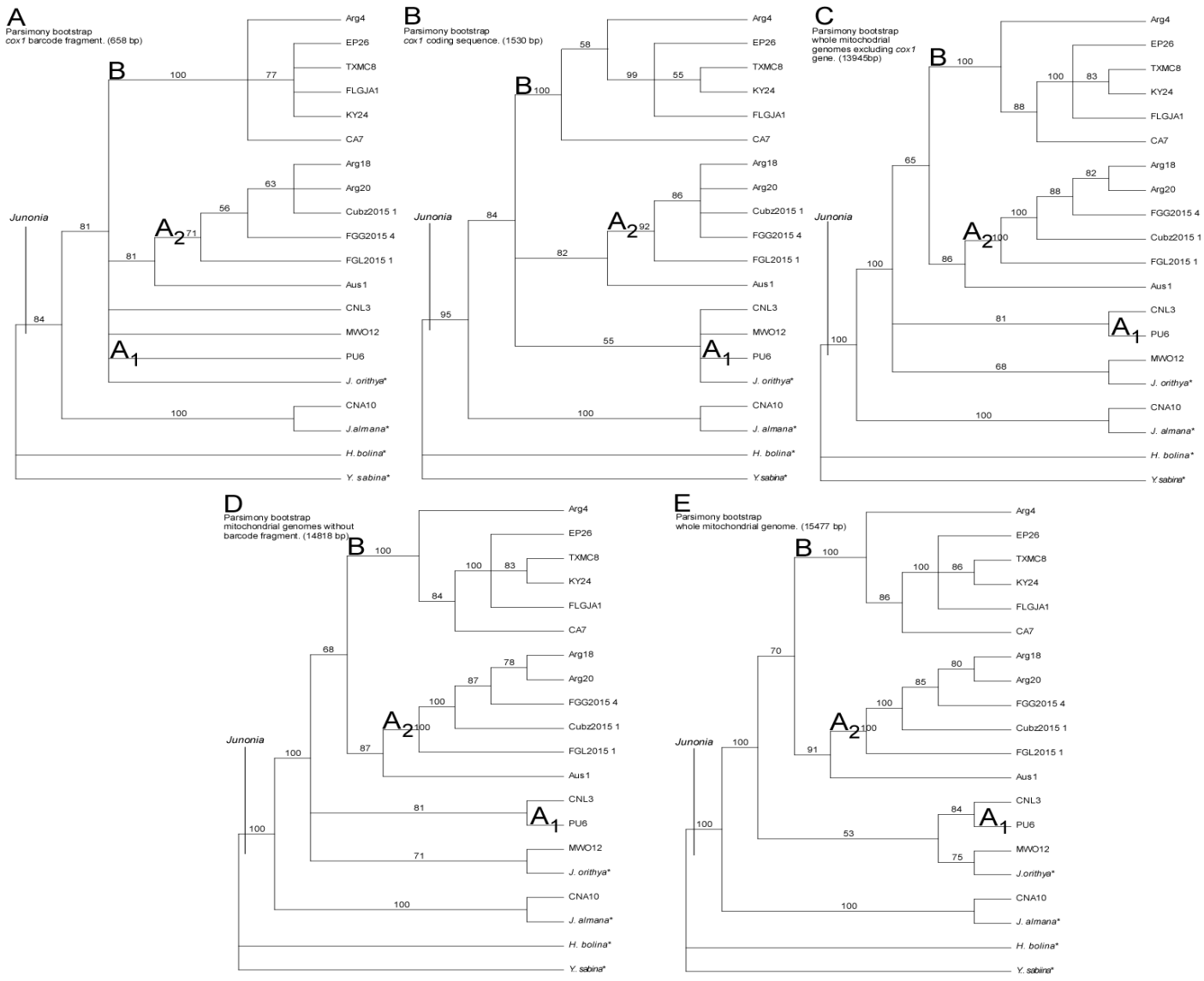

A
Parsimony bootstrap
cox1 barcode fragment. (658 bp)
B
A2
A1
Junonia
Arg4
EP26
TXMC8
FLGJA1
KY24
CA7
Arg18
Arg20
Cubz2015 1
FGG2015 4
FGL2015 1
Aus1
CNL3
MWO12
PU6
J. orithya*
CNA10
J.almana*
H. bolina*
Y. sabina*
B
Parsimony bootstrap
cox1 coding sequence. (1530 bp)
C
Parsimony bootstrap
whole mitochodrial
genomes excluding cox1
gene. (13945bp)
D
Parsimony bootstrap
mitochondrial genomes without
barcode fragment. (14818 bp)
E
Parsimony bootstrap
whole mitochondrial genome. (15477 bp)

Fig. 3-4. Maximum likelihood bootstrap analysis of genome data divided into subsets. Panel A: maximum likelihood bootstrap of *cox1* barcode fragment. Panel B: maximum likelihood bootstrap of mitochondrial genomes excluding *cox1* barcode fragment. Panel C: maximum likelihood bootstrap of whole mitochondrial genomes. Specimens denoted with an * are reference sequences downloaded from GenBank. *J. orithya*: accession number NC_022697. *J. almana*: accession number NC_024407. *H. bolina*: accession number NC_026072. *Y. sabina*: accession number NC_024403.

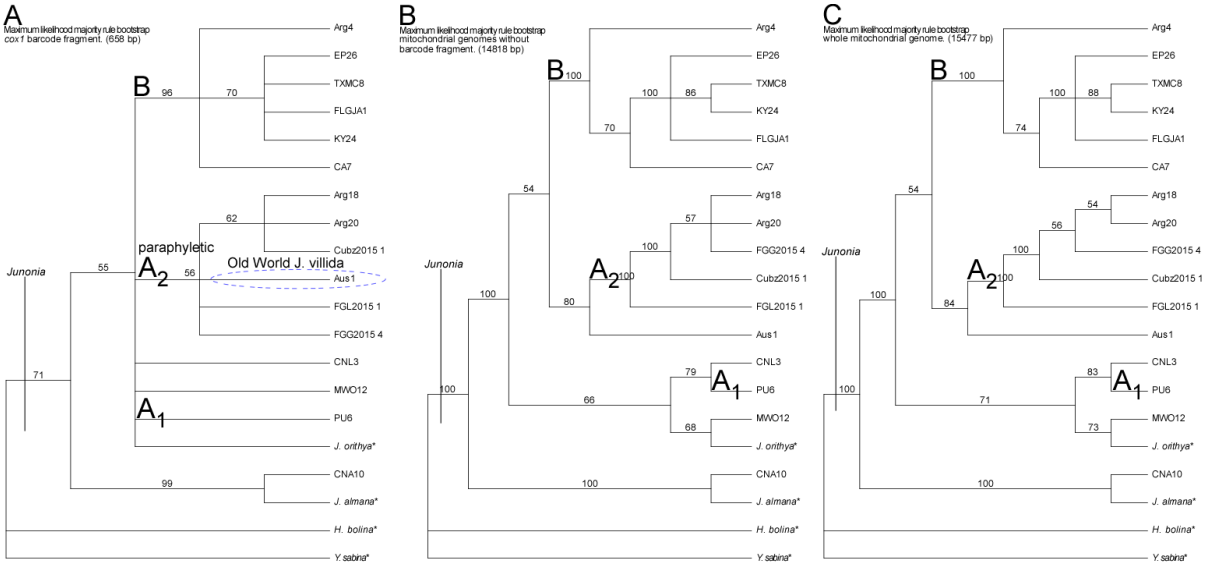


Table 3-5. Divergence times for clades of interest.

Comparison	**Haplotypes**	**% identity**	**% divergence**	**Divergence time (mya±95% confidence)**
J. orithya madagascariensis (MWO12) vs. *J. orithya orithya* reference	African vs. Asian subspecies of *J. orithya*	96.2	3.8	1.72±0.33
J. coenia (TXMC8) vs. *J. coenia grisea* (Ca7)	B diversification	97.5	2.5	1.18±0.29
J. vestina (PU6) vs. *J. lemonias* (CNL3)	A1 vs. *J. lemonias*	96.5	3.5	1.61±0.32
J. villida (AUS1) vs. *J. litoralis* (FGL2015.1)	A2 vs. *J. villida*	96.5	3.5	1.58±0.32
J. villida (AUS1) vs. *J. sp. flirtea* (Arg20)	A2 vs. *J. villida*	96.7	3.3	1.52±0.31
J. sp. hilaris (Arg4) vs. *J. coenia* (TXMC8)	B diversification	97.4	2.6	1.19±0.29
J. villida (AUS1) vs. *J. sp. hilaris* (Arg4)	B vs. *J. villida* and A2 lineage	96.1	3.9	1.79±0.34
J. villida (AUS1) vs. *J. coenia* (TXMC8)	B vs. *J. villida* and A2 lineage	94.9	5.1	2.31±0.42
J. sp. flirtea (Arg18) vs. *J. litoralis* (FGL2015.1)	A2 diversification	98.0	2.0	0.96±0.29

Discussion

Reconstructing the Evolutionary History of Junonia

The results of the phylogenetic analysis of the complete mitochondrial genomes (Fig. 3-2) are not consistent with the single invasion event hypothesis that has been the predominant interpretation of New World *Junonia* diversity (Figs. 1-1, 4-1) (Forbes, 1928, Eliot, 1946, Eliot, 1947, Forbes, 1947, Eliot, 1949, Kodandaramaiah & Wahlberg, 2007, Neild, 2008,

Kodandaramaiah, 2009, Pfeiler et al., 2012). The different New World *Junonia* haplotypes do not form a monophyletic group with each other to the exclusion of all Old World species (consistent with the predictions of hypothesis 1B), indicating that it is likely that multiple invasion events have occurred as hypothesized by Gemmell and Marcus (2015) (Gemmell & Marcus, 2015). Haplotype group A2 is more closely related to *J. villida* from Australia than it is to members of haplotype group B. This result echoes the haplotype network analysis conducted by Gemmell and Marcus (2015) and suggests that haplotype group A2 has more recent Asian ancestry than haplotype group B. Similarly, haplotype A1, found in the high elevation Andean species *J. vestina*, is the sister clade to *J. lemonias* from China with 90/83 maximum likelihood/parsimony bootstrap support, suggesting that this lineage also has independent Asian ancestry from the other clades. I have found no evidence in the scientific literature of *Junonia lemonias* ever being suggested as a close relative of any New World *Junonia* species. The sister clade to haplotype group B is the clade that contains *J. villida* and haplotype group A2. This suggests that the same lineage that gave rise to the clade containing *J. villida* and haplotype group was also the progenitor of New World haplotype group B. Consequently, while my data is not completely consistent with any of the morphology-based hypotheses for the origin of the New World *Junonia* taxa, it is closest to the hypothesis of Forbes (1928, 1947) (Hypothesis 2A) which predicted that *J. villida* was the most closely related Old World taxon to the New World *Junonia* (Fig. 1-1A).

These findings are consistent with the multiple invasion hypothesis proposed by Gemmell and Marcus (2015), though those authors only suggested two colonization events of the New World, rather than the 3 invasion events that appear to have taken place (Hypothesis C). The contrast between Fig. 3-1 and Fig. 3-2 is remarkable: *cox1* barcodes (658 nucleotides each)

show New World monophyly with weak bootstrap support (Fig. 3-1), while full length mitochondrial sequences (approximately 15,200 nucleotides each) resolve with moderate to high bootstrap support a paraphyletic relationship of the different New World haplotype groups to each other and to Old World *Junonia* species. Phylogenetic analysis of just the barcode fragments from within the complete mitochondrial genomes analyzed for this study (Fig. 3-3, Fig 3-4) suggests a lack of monophyly (with modest bootstrap support) within the New World Junonia, a result that differs from all prior studies of *cox1* barcodes in the New World *Junonia* (Brévignon & Brévignon, 2012, Pfeiler et al., 2012, Borchers & Marcus, 2014, Gemmell et al., 2014, Gemmell & Marcus, 2015). Modest sequence variation within *J. vestina* and *J. villida* (different specimens were used for the current work from the same collection localities as the specimens used in prior work), and the limited number of informative sites within a *cox1* barcode were sufficient to cause these differences. Phylogenetic analysis of larger portions of the mitochondrial genome also support a lack of monophyly of the New World Junonia (Fig. 3-3, Fig 3-4). Given the larger data set and the stronger bootstrap support for almost all nodes, the complete mitochondrial genome sequences produced a more robust phylogenetic hypothesis for the New World *Junonia*.

In my experiments, it cost approximately $315 to sequence a complete 15.2 kb butterfly mitochondrial genome, as compared to $13 to sequence a 658 bp butterfly mitochondrial *cox1* barcode. However, many research groups, including the Marcus laboratory, are developing techniques by which the simultaneous multiplex sequencing of many mitochondrial genomes within the same pooled DNA library is possible (Gillett et al., 2014, Timmermans et al., 2014), dramatically reducing the cost to approximately $30 per complete 15.2 kb mitochondrial genome. This is just over twice the price of sequencing a DNA barcode, but recovers 23 times as

much sequence data. The phylogenetic data sets compiled from whole mitochondrial genomes contain many more informative sites and therefore can produce better resolved phylogenetic trees than barcode sequences alone.

Based on the calculated divergence times (Table 3-5), haplotype B and haplotype A2 lineages diverged from each other between 1.79+/-0.34 and 2.31+/-0.42 million years ago , although this divergence may have occurred before either lineage arrived in the New World. This is similar to the 1.72+/-0.33 million year estimated divergence time between the Asian (*J. orithya orithya*) and African (*J. orithya madagascariensis*) subspecies of *J. orithya.* By approximately 1.18+/-0.29 to 1.19+/-0.29 million years ago, haplotype B was diversifying within the Western Hemisphere. Haplotype A2 likely diverged from *J. villida* from Australia between 1.52+/-0.31 to 1.59+/-0.32 million years ago. However, because *J. villida* occurs on many islands in the South Pacific that are geographically closer to the Americas (Vane-Wright & Tennent, 2011), Australian *J. villida* populations are unlikely to be the source of the individuals that colonized the New World. Examining additional *J. villida* mitochondrial genomes from other populations (such as the islands of French Polynesia in the Pacific) will help define the amount of intraspecific variation that occurs in this species and may help narrow the time frame during which the New World invasion took place. By approximately 0.96+/-0.29 million years ago, haplotype group A2 was diversifying within the New World. Haplotype group A1 that thus far only occurs in *J. vestina* diverged from *J. lemonias* approximately 1.60+/-0.32 million years ago, which is a maximum estimate for when haplotype A1 arrived in the New World.

The current geographic distributions of *J. lemonias* (Old World sister taxon to *J. vestina* and haplotype group A1) and *J. villida* (Old World sister taxon to haplotype group A2) in Southeast Asia and the Indo-Pacific are such that the easternmost extent of *J. lemonias* is in the

Philippines (17,500 km to the nearest New World landfall) (McCullagh & Marcus, 2015), and the easternmost extent of *J. villida* is in French Polynesia (8000 km to the nearest New World landfall) (Vane-Wright & Tennent, 2011). Such extreme distances suggest that favorable winds may have been essential for successful colonization of the New World. Consistent favorable Easterly winds in the South Pacific occur mostly south of the Tropic of Capricorn (23.5 degrees S latitude), but tend to turn northwards along the South American coast (Murck et al., 1995).The Altacama desert of Chile (Clarke, 2006) is the driest non-polar desert on earth, and is far older than any of the *Junonia* dispersal events currently under discussion, and stretches from 18 to 30 degrees S latitude. The Altacama may have prevented effective colonization in the southern parts of the continent. *Junonia vestina* and its associated haplotype group A1 are distributed in the Andes from northern Chile to southern Colombia, while haplotype group A2 is at its highest frequency (nearly 100%) in the lowlands to the west of the Andes in Peru and Ecuador (Gemmell & Marcus, 2015, M. Peters and J. M. Marcus, unpublished data). Collectively this suggests that the landfall site for both these genotypes in the Western Hemisphere was probably the Pacific Coast of South America, perhaps in present-day Peru or northern Chile. This is far enough south that dispersal could have been assisted by prevailing Easterly winds, but north of the most arid portions of the Altacama desert. Haplotype group B is more broadly distributed (McCullagh & Marcus, 2015), and may have been displaced by the arrival of the A1 and A2 haplotype groups in South America, so it is more difficult to infer where haplotype group B may have first made landfall in the New World.

In *Junonia*, in spite of the possible hybridization during its diversification in the New World, the use of full mitochondrial genome sequences clarifies the relationships between the New World haplotypes and those found in Old World species (Fig 3-2). In addition, the amount

of phylogenetic information in these mitochondrial genomes suggests that they can be used effectively for further exploration of relationships of genus *Junonia* with other Lepidopteran groups (Chapter 2, (McCullagh & Marcus, 2015)). Phylogenetic relationships among the Old World *Junonia* species, which are apparently not prone to hybridization, might also be clarified by analysis of complete mitochondrial genome sequences. However, within the New World *Junonia,* many of the relationships among mitochondrial haplotypes shown in the tree may not reflect the evolutionary history of the nuclear genomes of the various *Junonia* species, because the phenomena of hybridization and organelle capture may be obscuring these relationships (Mallet, 2007). Some New World *Junonia* species relationships have been suggested based on allelic variation at anonymous nuclear loci (Gemmell et al., 2014), but a phylogenomics approach using nuclear DNA sequences may be necessary to complete our understanding of the species-level phylogeny of this group. Other members of the Marcus lab are in the process of compiling a phylogenetic data set of this type.

Implications for DNA Barcoding

Although *cox1* barcodes are effective for identifying species for many of groups of animals (Hubert et al., 2008, Tavares & Baker, 2008), they are not informative in some groups due to recent divergence (and failure of coalescence), intraspecific selection for haplotype diversity, or hybridization (and organelle capture) (Avise & Saunders, 1984, Bernatchez, 1995, Mallet, 2005, Halbert & Derr, 2007, Mallet, 2007, Good, 2008, Schmidt & Sperling, 2008, Gemmell & Marcus, 2015). *Junonia* is clearly among these groups, with mitochondrial haplotype groups A2 and B, which differ by 3.9% *cox1* sequence divergence (Gemmell et al., 2014), both present in almost all New World *Junonia* species (Gemmell & Marcus, 2015).

However, *Junonia* also shows another limitation of the DNA barcoding approach. DNA barcoding studies have been criticized for the phylogenetic algorithms (especially neighbor-joining) that they frequently employ (Nixon & Wheeler, 1990, Brower, 2006) but this is easily corrected by analyzing the same data sets with other phylogenetic methods. Perhaps a more important limitation to phylogenetic analysis based on DNA barcode sequences is the limited number of species-specific informative positions within a short 658 bp stretch of barcode DNA (Spooner, 2009). If the barcode fragment contains insufficient information to produce a reliable phylogenetic hypothesis for the mitochondrial genome as a whole, and analysis of complete mitochondrial genomes reveals a substantially different and more robust tree, this suggests that the barcode fragment is an insufficient sample of informative sites within the mitochondrion (Funk & Omland, 2003, Schmidt & Sperling, 2008) and phylogenies based solely on the barcode fragment should be viewed with some skepticism.

Phylogenetic analysis of complete *Junonia* mitochondrial genome sequences shows that phylogenies based on mitochondrial *cox1* barcode sequences alone can be highly misleading. In particular, the placement of *J. vestina* haplotype A1 as the sister group to haplotype group A2 by several authors based on analysis of *cox1* barcodes, and the relationships of Old World *J. villida*, *J. orithya*, and *J. lemonias* to the New World *Junonia* (Pfeiler et al., 2012, Borchers & Marcus, 2014, Gemmell et al., 2014, Gemmell & Marcus, 2015) are due primarily to insufficient sampling of informative sites in the mitochondrial genome. Regardless of the phylogenetic analysis method used, molecular phylogenies based solely on short barcode sequences should be viewed with scepticism, particularly when phylogenetic reconstruction reveals limited bootstrap support for critical nodes. As it becomes more affordable to sequence complete mitochondrial genomes (by next generation sequencing or other techniques), it may be that these longer

sequences (roughly 23X larger than *cox1* barcodes), which are somewhat better for species diagnosis (though still vulnerable to producing misleading identifications due to recent divergence, lack of coalescence, or hybridization) and far more appropriate for phylogenetic analysis, will become "Next generation Barcodes" and replace the short barcoding fragments currently in use.

Acknowledgements

I thank Christian Brévignon, Carlos Cruz, Bob Dowell, and Ed Pfeiler for generously providing specimens for this project. Debbie Tsuyuki at the Next Generation Sequencing Platform facility at the Manitoba Institute of Child Health helped prepare samples and conduct the next generation DNA sequencing runs.

Literature Cited

Austin, G.T. and Emmel, J.F., 1998. New subspecies of butterflies (Lepidoptera) from Nevada and California. Systematics of western North American butterflies. (ed. T.C. Emmels), pp. 501-522. Marioposa Press, Gainesville, FL.

Avise, J.C. and Saunders, N.C., 1984. Hybridization and introgression among species of Sunfish (*Lepomis*) - Analysis by mitochondrial-DNA and allozyme markers. Genetics 108, 237-255.

Bernatchez, L.G., H. Wilson, C.C. Danzmann, R.G., 1995. Introgression and fixation of Acrtic char (*Salvelinus alpinus*) mitochondrial genome in an allopatric population of brook trout (*Salvelinus fontinalis*). Can. J. Fish. Aquat. Sci. 52, 179-185.

Borchers, T.E. and Marcus, J.M., 2014. Genetic population structure of buckeye butterflies (*Junonia*) from Argentina. Syst. Ent. 39, 242-255. doi: 10.1111/syen.12053.

Brévignon, C., 2004. Description de deux nouvelle sous-espèces Guadeloupeennes du genere *Junonia* Hübner, 1819 (Lepidoptera, Nymphalidae, Nymphalinae). Lambillionea CIV, 72-80.

Brévignon, C., 2009. Nouvelles observations sur le genre *Junonia* en Guyane Française. (Lepidoptera: Nymphalidae) Première Partie. Lambillionea CIX, 3-7.

Brévignon, L. and Brévignon, C., 2012. Nouvelles observations sur le genre *Junonia* en Guyane Française. (Lepidoptera: Nymphalidae) (Seconde partie). Lépidoptères de Guyane 7, 8-35.

Brower, A.V.Z., 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proc. Nat. Acad. Sci. USA 91, 6491-6495.

Brower, A.V.Z., 2006. Problems with DNA barcodes for species delimitation: 'Ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). Syst. Biodivers. 4, 127-132.
Burns, J.M., Janzen, D.H., Hajibabaei, M., Hallwachs, W. and Hebert, P.D.N., 2007. DNA barcodes of closely related (but morphologically and ecologically distinct) species of skipper butterflies (Hesperiidae) can differ by only one to three nucleotides. J. Lepid. Soc. 61, 138-153.
Clarke, J.D.A., 2006. Antiquity of aridity in the Chilean Atacama Desert. Geomorphology 73, 101-114.
Corbet, A.S., 1948. Papers on Malaysian Rhopalocera. V. The conspecificity of the American *Precis lavinia* (Cramer) with the Oriental *Precis orithya* (Cramer). Entomologist 81, 54-56.
Darriba, D., Taboada, G.L., Doallo, R. and Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Meth. 9, 772-772.
Dupuis, J.R. and Sperling, F.A.H., 2015. Repeated reticulate evolution in North American *Papilio machaon* group swallowtail butterflies. PLoS ONE 10, e0141882.
Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V.Z., Mallet, J. and Jiggins, C.D., 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proc. R. Soc. B: Biol. Sci. 274, 2881-2889.
Eliot, N., 1946. Continental drift and *Precis lavinia.* Entomologist 79, 225-228.
Eliot, N., 1947. More on continental drift, *Precis lavinia* Hb. and *P. villida* F. Entomologist 80, 230-234.
Eliot, N., 1949. The significance of wing pattern in *Precis* (Nymphalidae) in America. Entomologist 82, 176-184.
Forbes, W.T.M., 1928. Variation in *Junonia lavinia* (Lepidoptera, Nymphalidae). J. N.Y. Ent. Soc. 36, 306-321.
Forbes, W.T.M., 1947. Buckeyes and Wegener. Entomologist 80, 56-58.
Funk, D.J. and Omland, K.E., 2003. Species-level paraphyly and polyphyly: frequncy, causes, and consequences, with insights from animal mitochondrial DNA. Ann. Rev. Ecol. Evol. Syst. 34, 397-423.
Gaikwad, S.S., Ghate, H.V., Ghaskadbi, S.S., Patole, M.S. and Shouche, Y.S., 2012. DNA barcoding of nymphalid butterflies (Nymphalidae: Lepidoptera) from Western Ghats of India. Mol. Biol. Rep. 39, 2375-2383.
Gemmell, A.P., Borchers, T.E. and Marcus, J.M., 2014. Genetic Population Structure of Buckeye Butterflies (*Junonia*) from French Guiana, Martinique, and Guadeloupe. Psyche 2014, 1-21. doi: 10.1155/2014/897596.
Gemmell, A.P. and Marcus, J.M., 2015. A tale of two haplotype groups: The origin and distribution of divergent New World *Junonia COI* haplotypes. Syst. Ent. 40, 532-546. doi: 10.1111/syen.12120.
Gillett, C.P.D.T., Crampton-Platt, A., Timmermans, M.J.T.N., Jordal, B.H., Emerson, B.C. and Vogler, A.P., 2014. Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). Mol. Biol. Evol. 31, 2223-2237.
Good, J.M., Hird, S., Reid, N., Demboski, J.R., Steppan, S.J., Martin-Nimss, T.R., & Sullivan, J. , 2008. Ancient hybridization and mitochondrial capture between two species of chipmunks. Mol. Ecol. 17, 1313-1327.

Hafernik, J.E., 1982. Phenetics and ecology of hybridization in buckeye butterflies (Lepidoptera: Nymphalidae). Univ. Cal. Pub. Entomol. 96, 1-109.
Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W. and Hebert, P.D.N., 2006. DNA barcodes distinguish species of tropical Lepidoptera. Proc. Nat. Acad. Sci. USA 103, 968-971.
Halbert, N.D. and Derr, J.N., 2007. A comprehensive evaluation of cattle introgression into US federal bison herds. J. Heredity 98, 1-12.
Hebert, P.D.N., Cywinska, A., Ball, S.L. and deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B. 270, 313-321.
Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H. and Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Nat. Acad. Sci. USA 101, 14812-14817.
Hillis, D.M. and Bull, J.J., 1993. An empirical-test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182-192.
Hubert, N., Hanner, R., Holm, E., Mandrak, N.E., Taylor, E., Burridge, M., Watkinson, D., Dumont, P., Curry, A., Bentzen, P., Zhang, J., April, J. and Bernatchez, L., 2008. Identifying Canadian freshwater fishes through DNA barcodes. PLOS One 3, e2490.
Huelsenbeck, J.P., Hillis, D.M. and Nielsen, R., 1996. A likelihood-ratio test of monophyly. Syst. Biol. 45, 546-558.
Huelsenbeck, J.P. and Rannala, B., 1997. Phylogenetic methods come of age: Testing hypotheses in an evolutionary context. Science 276, 227-232.
Igarishi, S. and Fukada, H., 1997. The Life Histories of Asian Butterflies, Volume 1, Tokai University Press.
Igarishi, S. and Fukada, H., 2000. The Life Histories of Asian Butterflies, Volume 2, Tokai University Press.
Janzen, D.H., Hajibabaei, M., Burns, J.M., Hallwachs, W., Remigio, E. and Hebert, P.D.N., 2005. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. Phil. Trans. Roy. Soc. B 360, 1835-1845. doi:10.1098/rstb.2005.1715.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647-1649.
Kodandaramaiah, U., 2009. Eyespot evolution: Phylogenetic insights from *Junonia* and related butterfly genera (Nymphalidae: Junoniini). Evol. Dev. 11, 489-497.
Kodandaramaiah, U., Simonsen, T.J., Bromilow, S., Wahlberg, N. and Sperling, F.A.H., 2013. Deceptive single-locus taxonomy and phylogeography: *Wolbachia*-associated divergence in mitochondrial DNA is not reflected in morphology and nuclear markers in a butterfly species. Ecol. Evol. 3, 5167– 5176. doi: 10.1002/ece3.886.
Kodandaramaiah, U. and Wahlberg, N., 2007. Out-of-Africa origin and dispersal-mediated diversification of the butterfly genus *Junonia* (Nymphalidae: Nymphalinae). J. Evol. Biol. 20, 2181-2191.
Larsen, T.B., 1991. The Butterflies of Kenya and their Natural History, Oxford University Press, Oxford.
Mallet, J., 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20, 229-237.
Mallet, J., 2007. Hybrid speciation. Nature 446, 279-283.

McCullagh, B.S. and Marcus, J.M., 2015. The complete mitochondrial genome of the lemon pansy *Junonia lemonias* (Lepidoptera: Nymphalidae: Nymphalinae). J. Asia Pac. Entomol. 18, 749-755.
Mecenero, S., Ball, J.B., Edge, D.A., Hamer, M.L., Henning, G.A., Kruger, M., Pringle, E.L., Terblanche, R.F. and Williams, M.C., 2013. Conservation Assessment of Butterflies of South Africa, Lesotho and Swaziland – Red List and Atlas, Saftronics and the Animal Demography Unit, University of Cape Town.
Minno, M.C. and Emmel, T.C., 1993. Butterflies of the Florida Keys, Mariposa Press, Scientific Publishers, Gainesville, Florida.
Munroe, E.G., 1951. The genus *Junonia* in the West Indies (Lepidoptera, Nymphalidae). Am. Mus. Novit. 1498, 1-16.
Murck, B.W., Skinner, B.J. and Porter, S.C., 1995. Environmental Geology, New York.
Neild, A.F.E., 2008. The Butterflies of Venezuela, Part 2: Nymphalidae II (Acraeinae, Libytheinae, Nymphalinae, Ithomiinae, Morphinae), Meridian, London.
Nixon, K.C. and Wheeler, Q.D., 1990. An amplification of the phylogenetic species concept. Cladistics 6, 211-223.
Paulsen, S.M., 1996. Quantitative genetics of the wing color pattern in the buckeye butterfly (*Precis coenia* and *Precis evarete*): evidence against the constancy of G. Evolution 50, 2360-2372.
Pfeiler, E., Johnson, S. and Markow, T.A., 2012. DNA barcodes and insights into the relationships and systematics of buckeye butterflies (Nymphalidae: Nymphalinae: *Junonia*) from the Americas. J. Lepid. Soc. 66, 185-198.
Ratnasingham, S. and Hebert, P.D.N., 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). Mol. Ecol. Notes 7, 355-364.
Rice, P., Longden, I. and Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics 16, 276-277.
Schmidt, B.C. and Sperling, F.A.H., 2008. Widespread decoupling of mtDNA variation and species integrity in *Grammia* tiger moths (Lepidoptera: Noctuidae). Syst. Entomol. 33, 613-634.
Shi, Q.-H., Sun, X.-Y., Wang, Y.-L., Hao, J.-S. and Yang, Q., 2015. Morphological Characters Are Compatible with Mitogenomic Data in Resolving the Phylogeny of Nymphalid Butterflies (Lepidoptera: Papilionoidea: Nymphalidae). PLoS ONE 10.
Shi, Q., Huang, D., Wang, Y. and Hao, J., 2013. The complete mitochondrial genome of Blue Pansy, *Junonia orithya* (Lepidoptera: Nymphalidae: Nymphalinae). Mitochondrial DNA 0, 1-2. doi: 10.3109/19401736.19402013.19823182.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D. and Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7.
Spooner, D.M., 2009. DNA barcoding will frequently fail in complicated groups: an example in wild potatoes. Am. J. Bot. 96, 1177-1189.
Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts, USA.
Tavares, E.S. and Baker, A.J., 2008. Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. BMC Evol. Biol. 8, 81 doi:10.1186/1471-2148-8-81.

Tilden, J.W., 1970. Comments on the Nearctic members of the genus *Precis* Huebner. J. Res. Lep. 9, 101-108.
Timmermans, M.J.T.N., Lees, D.C. and Simonsen, T.J., 2014. Towards a mitogenomic phylogeny of Lepidoptera. Mol. Phylogenet. Evol. 79, 169-178.
Turner, T.W. and Parnell, J.R., 1985. The identification of two species of *Junonia* Hübner (Lepidoptera: Nymphalidae): *J. evarete* and *J. genoveva* in Jamaica. J. Res. Lep. 24, 142-153.
Vane-Wright, R.I. and Tennent, W.J., 2011. Colour and size variation in *Junonia villida* (Lepidoptera, Nymphalidae): subspecies or phenotypic plasticity? Syst. and Biodivers. 9, 289-305.
Wahlberg, N., Weingartner, E., Warren, A. and Nylin, S., 2009. Timing major conflict between mitochondrial and nuclear genes in species relationships of *Polygonia* butterflies (Nymphalidae: Nymphalini). BMC Evol. Biol. 9, 92.
Wu, L.-W., Lin, L.-H., Lees, D. and Hsu, Y.-F., 2014. Mitogenomic sequences effectively recover relationships within brush-footed butterflies (Lepidoptera: Nymphalidae). BMC Genomics 15, 468 doi:10.1186/1471-2164-15-468.
Zander, R.H., 2004. Minimal values for reliability of bootstrap and jackknife proportions, decay index, and Bayesian posterior probability. PhyloInformatics 2, 1-13.

Chapter 4: Into thin air: Testing hypotheses for high altitude mitochondrial genome evolution in *Junonia* butterflies

Abstract

The effects of high altitude conditions on organisms living at these conditions is well studied in mammals but less so in other groups. *Junonia vestina*, a butterfly found in the Andes Mountains, is a candidate for examining potential high altitude adaptations in the mitochondrial genome. I compared the mitochondrial genomes of *J. vestina* to low elevation sister taxon *J. lemonias* and outgroup *J. orithya madagascariensis* to detect evidence of high altitude adaptation. The mitochondrial genome of *J. vestina* has an elevated %AT content compared to low elevation *Junonia* species. Compared to *J. lemonias, J. vestina* shows elevated non-synonymous mutation rates in the mitochondrial protein coding genes *atp6, atp8, cob, cox1, nad1,* and *nad2.* Many of these mutations are in transmembrane domains that line proton channels through the inner mitochondrial membrane. Mitochondrial genes identified in *J. vestina* were compared with mitochondrial genomes from other high elevation species. The *atp6, atp8, cob,* and *nad2* loci were most frequently implicated as having undergone adaptation to high elevation in high altitude organisms with sequenced mitochondrial genomes, but the strongest signals for adaptation to high elevation in *J. vestina* were at the *cox1* and *nad1* genes. Natural non-synonymous variation also occurs at many of the same sites in the human mitochondrion and some are associated with pathological conditions. One mutation in *cox1* (position G136S in *J. vestina*, P139S in *H. sapiens*) is near the active site and may facilitate pre-loading the site with oxygen molecules when they occur at low concentration.

Keywords: *Junonia vestina*, high altitude adaptation, ectothermy, mitochondrial genome

Introduction

The mitochondrion is an essential structure for respiration in the cells of all aerobic eukaryotes. Two of the three major components of eukaryotic respiratory metabolism, the Krebs cycle and the electron transport chain, occur within this organelle. Only glycolysis, the first step of respiration, occurs outside the mitochondrion in the cytoplasm (Berg et al. 2007). The mitochondrion possesses its own genome, separate from that of the nucleus, which encodes many of the enzymes responsible for catalyzing the reactions of the later stages of aerobic respiration (Berg et al. 2007). While mitochondrial genes perform essential functions in virtually all aerobic environments (Irwin, 2012), environments that place extreme demands on the apparatus of respiration (e.g. high or low temperatures or low oxygen concentrations) would be expected to exert strong selection for alleles encoding enzymatically-efficient gene products under those conditions.

High elevation habitats are a good example of an extreme environment for mitochondrial function. Adaptation to high elevation environments has been well studied in mammals, particularly in humans (Luo et al., 2013). The extreme conditions at high elevation, including low temperature and low atmospheric oxygen concentration may necessitate metabolic adaptations for respiratory function. Much research has been conducted on human populations living at high elevations such as the Sherpas and Tibetans, with a subset of this research focused on mitochondrial genomes (Torroni et al., 2005, Gu et al., 2012, Kang et al., 2013). Within the Sherpa population, mitochondrial *nad1* was identified as having several mutations that may have been positively selected in this population as an adaptation for high elevation conditions (Kang et al., 2013). In the ethnic Tibetan population, compared to the sympatric Han Chinese population which has lived on the Tibetan plateau for a much shorter period of time, higher rates of non-

synonymous and synonymous substitution were found in *atp6, atp8,* and *cytochrome oxidase b* (*cob*) (Gu et al., 2012).

Additional research on high elevation adaptation has been conducted on a variety of mammals, such as the Tibetan antelope (*Pantholops hodgonsii*) (Xu et al., 2005), the Tibetan horse (*Equus caballus)* (Xu et al., 2007), the Tibetan wild ass (*Equus kiang*) (Luo et al., 2012), the plateau pika (*Ochotona curzoniae*) (Luo et al., 2008), the Chinese snub nose monkey (*Rhinopithecus roxellana*) (Yu et al., 2011), and multiple high elevation-adapted species of the Mammalian tribe Caprini (which includes sheep, goats, and their relatives) (Hassanin et al., 2009). In the Tibetan antelope, several *cox1* mutations were found with a potential link to high altitude adaptations (Xu et al., 2005). In Tibetan horses, from three different regions of Tibet, a high rate of non-synonymous substitution in *nad6* was identified in all three of the sequenced mitochondrial genomes (Xu et al., 2007). The plateau pika, which occurs from a range of 3000-6000 m above sea level in Tibet, has mutations in the transmembrane region of *cox1* along with a non-synonymous change in *cox2* (Luo et al., 2008). In a study of two genes (*atp6, cox1*) of the Tibetan chicken (*Gallus gallus*) indications of high altitude adaptation were found in *atp6* (Zhao et al., 2015). Studies in these diverse groups of organisms suggest that there are some features of mitochondrial genomes of species living at high elevations that seem to be shared across different mammalian species and groups.

Many different mitochondrial genes have been implicated within these species as potentially having functions related to high altitude adaptation in mammals, including *atp6, atp8, cox1, cox2, cob, nad1, nad2, nad3, nad4, nad5,* and *nad6* (Luo et al., 2013). Of these genes, non-synonymous sequence evolution in *atp6, atp8, cox1, cob, nad2,* and *nad6* have most frequently been associated with adaptive evolution to high altitude (Table 4-1). In addition, in some cases,

high-altitude species show elevated proportions of A and T nucleotides in the mitochondrial genome when compared to closely related species living at lower altitudes (Hassanin et al., 2009). This appears to be caused high rates of deamination of cytosine bases on the heavy strand of the mitochondrial genome, causing C to T transitions (Hassanin et al., 2009).

Two hypotheses have been suggested to explain the pattern of AT-richness at high altitude: the occurrence of severe oxidative stress and higher rates of metabolism (Hassanin et al., 2009). The severe oxidative stress hypotheses suggests that mitochondria in organisms that live at high altitude experience low atmospheric oxygen concentrations, increased UV radiation, and low temperatures, each of which causes an increase in the frequency of reactive oxygen species (Hassanin et al., 2009). Synergistically, these environmental factors can cause severe oxidative stress which has been shown to increase rates of cytosine deamination (Kow, 2002) and an elevated mutation rate at high altitudes.

The high altitude high metabolic rate hypothesis suggests that mammals might need higher metabolic rates to maintain endothermy at the low temperatures found at high altitude (Hassanin et al., 2009). Maintaining higher metabolic rates can require larger numbers of mitochondria and higher rates of DNA synthesis to produce copies of the mitochondrial genome (Martin & Palumbi, 1993). Mutation rates are positively correlated with rates of DNA synthesis so an increase in metabolic rate due to increased environmental pressure to maintain body temperatures can cause an increase in the rate of DNA sequence evolution (Martin & Palumbi, 1993). Additionally, the presence of additional mitochondria alone might increase the incidence of reactive oxygen species and further increase the mutation rate in mitochondrial DNA (Martin & Palumbi, 1993).

Table 4-1. Suspected adaptive changes in protein coding genes in high-altitude organisms. Genes identified as candidates for high elevation adaptation after analysis of sequenced mitochondrial genomes are indicated by x.

Protein Coding Genes	*atp6*	*atp8*	*cox1*	*cox2*	*cox3*	*cob*	*nad1*	*nad2*	*nad3*	*nad4*	*nad4L*	*nad5*	*nad6*	Reference
Species														
Junonia vestina	x	x	x			x	x	x						This study
Homo sapiens (Tibet)	x	x				x								Gu, 2012
Homo sapiens (Sherpa)							x							Kang, 2013
Ochotona curzoniae (Plateau pika)			x	x		x		x	x			x	x	Luo, 2008
Triplophysa stoliczkae (Tibetan stone loach)														Sequenced by Li et al., 2013. Further analysis in this study
Equus caballus (Tibetan horses)													x	Xu et al., 2007
Artemia tibetiana (brine shrimp)	x	x		x	x									Zhang et al., 2013
Pantholops hodgsonii (Tibetan antelope)			x											Xu et al., 2005
Schizothoracine fishes	x					x	x	x		x		x		Li et al., 2013b

Mammalian Tribe Caprini	x	x												Hassanin et al., 2009
Rhinopithecus roxellana (Chinese snub-nosed monkeys)								x					x	Yu et al., 2011
Equus kiang (Tibetan wild ass)	x	x		x		x		x		x		x		Luo et al., 2012
Total examples	6	5	3	3	1	5	3	5	1	2	0	3	3	
Studies examining each gene	12	12	12	12	12	12	12	12	12	12	12	12	12	

It is difficult to distinguish between these alternatives in endotherms such as mammals because at high altitudes these organisms simultaneously experience both high metabolic rates to maintain body temperature as well as low atmospheric oxygen concentrations and high UV exposure (Hassanin et al., 2009), making it difficult to determine what is driving AT-richness in the mitochondria of endotherms. However, the study of high elevation adapted ectothermic species may offer the possibility of distinguishing between these two alternatives. Because butterflies are ectothermic and have body temperatures that generally track environmental temperatures (Kleckova et al., 2014), any changes in AT-richness in the high-altitude butterfly mitochondria cannot be a consequence of higher metabolic rates required for endothermic temperature regulation, and thus may be attributable to the effects of low oxygen tension in the environment.

Prior to the current study, the only ectothermal high elevation species with sequenced mitochondrial genomes are a brine shrimp, *Artemia tibetiana* (Zhang et al., 2013), and 3 fish species, the Tibetan stone loach, *Triplophysa stoliczkae* (Cobitidae) (Li et al., 2013a), and 2 high elevation species of Schizothoracine fishes (Cyprinidae) (Li et al., 2013b). All are included in Table 4-1. *Artemia tibetiana* shows possible adaptive mutations in *atp6, atp8, cox2,* and *cox3* (Zhang et al., 2013). *Triplophysa stoliczkae* shows no non-synonymous changes from its sister taxon (a low elevation species), and presumably from their most recent common ancestor (Li et al., 2013a). The 2 high elevation Schizothoracine fishes show possible adaptive mutations in *atp6, cob, nad1, nad2, nad4,* and *nad5* (Li et al., 2013a). Some of these genes (*atp6, atp8,* and *cob*) have been identified as possible sites of adaptation in many earlier studies of high elevation mammals (see discussion above), while another (*nad4*) were only identified in a single study of Tibetan wild ass (Luo et al., 2012). The generality of any statements regarding similarities or

differences between high elevation endotherms and ectotherms requires additional sampling of ectotherms. For example, there are no previous reports of mitochondrial genomes sequences available from any high elevation insect species (which are typically ectotherms).

This makes the availability of a complete mitochondrial genome sequence from *Junonia vestina*, the Andean buckeye, a very useful contribution to high elevation mitochondrial adaptation research. *Junonia vestina* is a high elevation species found exclusively at elevations of 1900 m up to approximately 4000 m in the Andes Mountains of South America (Pfeiler et al., 2012). It possesses a unique mitochondrial haplotype (A1) that distinguishes it from all of its congeners (Pfeiler et al., 2012) (Table 4-2). Given that *J. vestina* is the only New World *Junonia* species that occurs at high altitude (Table 1-1, other species typically occur below (and often much below) 2000 m) (Neild, 2008, Gemmell & Marcus, 2015) and it is derived from a low elevation ancestor (Chapter 3), examining the mitochondrial genome of this species and comparing it to close relatives (inferred sister taxon *J. lemonias* and outgroup *J. orithya madagascariensis)* can be used to distinguish between the high metabolic rate and low oxygen tension hypotheses for high elevation AT-richness. In addition, it may be possible to detect candidate high altitude adaptations in protein coding genes in this ectothermic species and to compare them with suspected high altitude adaptations identified in humans and other mammals. This may provide insights into the likely functional effects of these mutations, with implications for reconstructing scenarios of high elevation mitochondrial evolution in butterflies

Since doing functional tests of genetic changes are extremely difficult within *Junonia* butterflies, human pathology associated with the same codons can be used to infer possible functional consequences of mutations at these codon sites within mitochondrial proteins with highly conserved roles in the electron transport chain. Thousands of human mitochondrial

genomes have been sequenced, and epidemiological studies have associated pathology with a subset of the observed human variation (Lott et al., 2002). Human pathology associated with mutations at a particular codon position suggests that this position could have a biochemically important role in the function of the associated gene product. Human and butterfly physiology are profoundly different in many ways, but the electron transport chain of oxidative respiration is fundamentally important to both, so mutant positions that alter mitochondrial gene product function in humans (in order to produce pathology) can suggest mutations in the same position in butterflies may also have some physiological effect (though it may be different in detail) in butterflies as well. In contrast, parallel non-synonymous mutations at homologous codon positions in humans and butterflies, which have no apparent phenotypic effect in humans, may also have negligible phenotypic effect in butterflies, especially if the amino acid substitutions are conservative. Due to the conservation of mitochondrial genome structure and function, comparing wildtype and mutant mitochondrial DNA sequences among and between taxonomically disparate organisms in similar environments is a common form of inference in studies of mitochondrial functional adaptation in order to identify DNA sequence variation that is likely to have functional consequences (e.g. (Balloux et al., 2009, Foote et al., 2010, Luo et al., 2013, Finch et al., 2014)).

Objectives:

1. To distinguish between the environmental and metabolic hypotheses for elevated mutation rates and AT richness in high elevation animals.

Hypothesis 1A: High elevation environmental conditions alone are sufficient to produce elevated mutation rates and AT richness in the mitochondria of organisms adapted to high altitude.

Predictions:

1. AT richness should be a common feature of the mitochondria of high elevation-adapted organisms.

2. High elevation ectotherms such as *J. vestina* butterflies should show similar patterns of AT richness to what has previously been observed in endothermic mammals.

Hypothesis 1B: High elevation environmental conditions alone are insufficient to produce elevated mutation rates and AT richness in the mitochondria of organisms adapted to high altitude and these patterns are only seen in endothermic species which are predicted to have elevated metabolic rates at high elevation.

Predictions:

1. AT richness should be a common feature of the mitochondria of high elevation-adapted organisms.

2. High elevation ectotherms such as *J. vestina* butterflies should not show similar patterns of AT richness to what has previously been observed in endothermic mammals.

2. To investigate the differences between high elevation *J. vestina*, low elevation sister taxon *J. lemonias*, and low elevation outgroup *J. orithya madagascariensis* in order to identify mitochondrial sequence divergence in protein coding genes that may be related to high altitude adaptation.

Hypothesis 2A: Adaptation to high elevation conditions is occurring in mitochondrial protein coding genes in *J. vestina*.

Predictions:

1. There are non-synonymous differences between *J. vestina* and low elevation outgroup *J. orithya madagascariensis*.

2. Some of these non-synonymous differences are autapomorphic to the *J. vestina* lineage and are not shared with low elevation sister taxon *J. lemonias*.

3. Genes undergoing adaptive selection in *J. vestina* may show a disproportionately high frequency of autapomorphic non-synonymous mutations relative to sister taxon *J. lemonias*. (If prediction 3 is correct, it strengthens support for this hypothesis, but if this prediction is incorrect, adaptive evolution could still be taking place).

Hypothesis 2B: Adaptation to high elevation conditions is not occurring in mitochondrial protein coding genes in *J. vestina*.

Predictions: (Any of the following would support this hypothesis)

1. There are no non-synonymous differences between *J. vestina* and low elevation outgroup *J. orithya madagascariensis*.

2. There are no autampomorphic non-synonymous differences between *J. vestina* and *J. lemonias.*

3. There are autampomorphic non-synonymous differences between *J. vestina* and *J. lemonias*, but they are amino acid changes in regions of the protein that do not affect protein function.

3. To determine if mitochondrial mutations related to high altitude adaptation in *J. vestina* also occur in other high elevation species, and whether there are any diseases associated at the same amino acid positions within the same genes in human populations. Pathology associated with specific amino acid positions in mitochondrial gene products can help associate biological function with specific amino acids in a peptide sequence. This can give clues to the possible functional consequences of mutations at the same position in high elevation butterflies, and help determine which mutated sites in *J. vestina* are perhaps most functionally significant.

Hypothesis 3A: High altitude conditions pose similar environmental challenges to all animals living at high elevation, selecting for parallel evolution of features of the mitochondrial genome to respond to these conditions.

Predictions:

1. The same genes will be important for high elevation adaptation in all high elevation animals.

2. Mutations at the same sequence positions within protein coding genes may be observed in multiple high elevation species.

3. These mutations may be occurring in the same sequence positions as mutations in human populations that are associated with visible phenotypes (such as pathology). The presence of phenotypic variation associated with genetic variation at a particular sequence position is an indication that the position is important to the function of the encoded gene product.

Hypothesis 3B: High altitude conditions pose different challenges to each species living at high elevation and adaptive evolution to high altitude does not produce patterns of parallel evolution in the mitochondrial genome.

Predictions:

1. Different genes will be important for high elevation adaptation in various high elevation animals.

2. Mutations at the same sequence positions within protein coding genes will generally not be observed in multiple high elevation species.

Methods

Complete mitochondrial genomes of 14 *Junonia* butterflies were sequenced and assembled as described in Chapter 3. Percent GC and AT content were calculated, and GC and AT skew was also determined (Lobry, 1996) (Table 4-2). In *J. vestina*, the protein coding mitochondrial genes most likely to be under selective pressure were identified by aligning the *J. vestina* mitochondrial genome sequence to the mitochondrial genome of its inferred low elevation sister species *J. lemonias* along with a near outgroup for these two taxa: *J. orithya madagascariensis*, an Old World *Junonia* species from Africa. The *J. orithya madasgascariensis* complete mitochondrial genome sequence was selected for use as an outgroup because it is one of two sequences (the other is *J. orithya orithya* from Asia (Shi et al., 2013)) that form a sister group to the *J. lemonias*-*J. vestina* clade, though with only moderate bootstrap support (Fig. 3-2). However, these relationships are more strongly supported than the results from any prior phylogenetic analyses, which fail to resolve any of these nodes with any statistical confidence (Kodandaramaiah & Wahlberg, 2007, Kodandaramaiah, 2009). I chose to use the *J. orithya madagascariensis* sequence because I obtained it myself and can vouch for the quality of the sequence reads and mitochondrial genome assembly (Chapter 3). However, if the *J. orithya orithya* sequence (Shi et al., 2013) or the reconstructed ancestor of the two *J. orithya* sequences

are used instead for the analyses presented here, there is no substantive change in any of the results or conclusions of this study (data not shown).

The alignments were performed using Clustal Omega (Sievers et al., 2011) using default parameters. All synonymous and non-synonymous nucleotide changes were identified and noted in both *J. vestina* and *J. lemonias* compared with *J. orithya madagascariensis*. Sequence selection was tested with *J. vestina, J. lemonias,* and *J. orithya madagascariensis* using Tajima's D test (Tajima, 1989) for selection in MEGA 6.06 (Tamura et al., 2013).

To look for shared gene changes in *J. vestina* compared to other high altitude adapted animals, genes of interest identified by other authors were compared to the *J. vestina* sequence (Table 4-1). For groups of organisms where mitochondrial genome sequences were published without interspecific comparative analyses (e.g. the loaches in genus *Triplophysa*), I downloaded the mitochondrial genome sequences from GenBank and aligned the protein-coding mitochondrial genes of the high elevation species to closely related species (usually congeners) that occur at lower elevations. For genes in which elevated rates of non-synonymous mutations were detected in *J. vestina* (Table 4-3), I determined if there were parallel non-synonymous changes in these genes between *J. vestina* and other high elevation species that had been studied previously.

Table 4-2. Old World and New World mitochondrial genomes sequenced.

Species	Voucher	Origin	% AT	Citation/Haplotype
Old World				
J. almana	CNA10	China	80.20	This study
J. almana	NC_024407.1	Japan	80.26	Wu et al., 2014
J. lemonias	CNL3	China	80.40	Chapter 2; McCullagh & Marcus, 2015
J. orithya	NC_022697.1	China	80.40	Shi et al., 2013
J. orithya madagascariensis	MWO12	Malawi	80.90	This study
J. villida	Aus1	Australia	80.60	This study
New World				
J. vestina	PU6	Peru	81.10	This study; A1
J. coenia	KY24	Kentucky, USA	80.60	This study; B
J. coenia	TXMC8	Texas, USA	80.74	This study; B
J. coenia grisea	CA7	California, USA	80.70	This study; B (California variant)
J. genoveva	FLGJa1	Florida, USA	80.70	This study; A2
J. genoveva	FGG2015.4	French Guiana	80.44	This study; A2
J. "genoveva" (mangrove feeder)	EP26	Mexico	80.88	This study; B
J. litoralis	FGL2015.1	French Guiana	80.75	This study; A2
J. sp. flirtea (light phenotype)	Arg18	Argentina	80.80	This study; A2
J. sp. flirtea (dark phenotype)	Arg20	Argentina	80.71	This study; A2
J. sp. hilaris	Arg4	Argentina	80.80	This study; B
J. zonalis	Cubz2015.1	Cuba	79.78	This study; A2

Table 4-3. Synonymous and non-synonymous changes in *J. vestina* and *J. lemonias* mitochondrial protein coding genes compared to *J. orithya madagascariensis.*

Compared to *J. orithya*	***J. lemonias***	***J. vestina***	**Shared NS sites**	**Unique NS sites *J. lemonias***	**Unique NS sites *J. vestina***
Gene	(NS/S)	(NS/S)			
atp6	1/17	2/17	1	0	1
atp8	1/2	2/2	1	0	1
cob	2/59	2/39	1	1	1
cox1	0/60	2/49	0	0	2
cox2	5/27	3/25	2	3	1
cox3	4/24	4/26	3	1	1
nad1	3/37	8/22	1	2	7
nad2	2/34	3/31	1	1	3
nad3	2/15	2/11	1	1	1
nad4	19/44	15/48	12	7	3
nad4L	15/8	15/8	14	1	1
nad5	14/45	9/39	6	8	2
nad6	6/14	4/8	3	3	1
				Total 28	Total 25

The synonymous and non-synonymous changes in *J. vestina* and *J. lemonias* were further analyzed by taking the number of non-synonymous changes and dividing them by the number of non-synonymous changes. This was done for each protein coding gene for both *J. vestina* and *J. lemonias*. The *J. vestina* ratio value for each protein coding gene was then divided by the ratio value of *J. lemonias* for each protein coding gene.

Finally, the genes with elevated non-synonymous sites in *J. vestina* were aligned to the same genes from a human (*Homo sapiens*) wild-type mitochondrial genome sequence (GenBank voucher XEB004 (Kang et al., 2013)). After generating the alignment in Clustal Omega, the *J.*

vestina non-synonymous positions were added to the sequence alignment to identify the homologous positions in the *H. sapiens* sequence. Using MITOMAP (Lott et al., 2002), these specific positions were checked against the indexed human mitochondrial genomes to determine if there were any pathological conditions associated with these codons (Table 4-4).

Table 4-4. Amino acid changes in *J. vestina* and *H. sapiens*. Associated pathology with *H. sapiens* mutations is listed.

Gene	**Amino acid (AA) position in *Junonia* (written as *J. orithya* AA, position number, *J. vestina* AA)**	**Homologous amino acid position in the *H. sapiens* gene (written as wildtype AA, position number, mutant AA)**	**Nucleotide change in *H. sapiens* (written as wildtype nucleotide, position number, mutant nucleotide)**	**Associated pathology**	**Reference**
ATP synthase subunit 6 (atp6)	S64T	T63A	A8713G	No associated condition	Lott et al., 2002, Tang et al., 2013
		T63I	C8714T	Various	
ATP synthase subunit 8 (atp8)	M35I	M42L	A8489T	No associated condition	Lott et al., 2002, Ban et al., 2008, Yao et al., 2008
		M42V	A8489G	No associated condition	
		M42T	T8490C	Multiple sclerosis	
Cytochrome B (cob)	L332M	L328P	T15729C	No associated condition	Lott et al., 2002
Cytochrome C oxidase subunit I (*cox1*)	G136S	P139S	C6318T	Intercellular calcium function changes	Kazuno et al., 2006
Cytochrome C oxidase subunit I	I466V	I469V	A7308G	No associated condition	Lott et al., 2002, Lehtonen et al., 2003, Puomila et al., 2007
				No associated condition	

(*cox1*)		I469L	A7308C		
		I469T	T7309C	familial sensorineural hearing impairment, Leber hereditary optic neuropathy (LHON)	
NADH dehydrogenase subunit 1 (*nad1*)	I9V	P2S	C3310T	Diabetes (type 2), hypertrophic cardiomyopathy	Hattori et al., 2003, Hattori et al., 2005, Chen et al., 2006
					Lott et al., 2002
		P2L	C3311T	No associated condition	
NADH dehydrogenase subunit 1 (*nad1*)	I11V	A4T	G3316A	Alzheimer's disease, deafness, diabetes (type 2), elevated blood pressure, LHON, megakaryoblastic leukaemic cells	Mostafaie et al., 2004, Tang et al., 2006, Piccoli et al., 2008, Lu et al., 2010, Qu et al., 2010, Liu et al., 2012, Kang et al., 2013
		A4V	C3317T	No associated condition	or Lott et al., 2002
NADH dehydrogenase subunit 1 (*nad1*)	GG312D	S312P	T4240C	No associated condition	Lott et al., 2002

Results

When compared to other sequenced *Junonia* mitochondrial genomes (Table 4-2), *J. vestina* had an AT content of 81.10%, which is at the upper bound of the 95% confidence limit of the mean AT richness (80.59%± 0.5%) for genus *Junonia*. While formal statistical tests are not possible between a single data point and a range, this suggests that the AT content of high altitude *J. vestina* may differ substantially when compared to its low elevation congeners. Although it would be beneficial to compare the composition of the *J. vestina* mitochondrial genome sequence with those from other high elevation butterfly (or insect) species, no such sequences have been published to date.

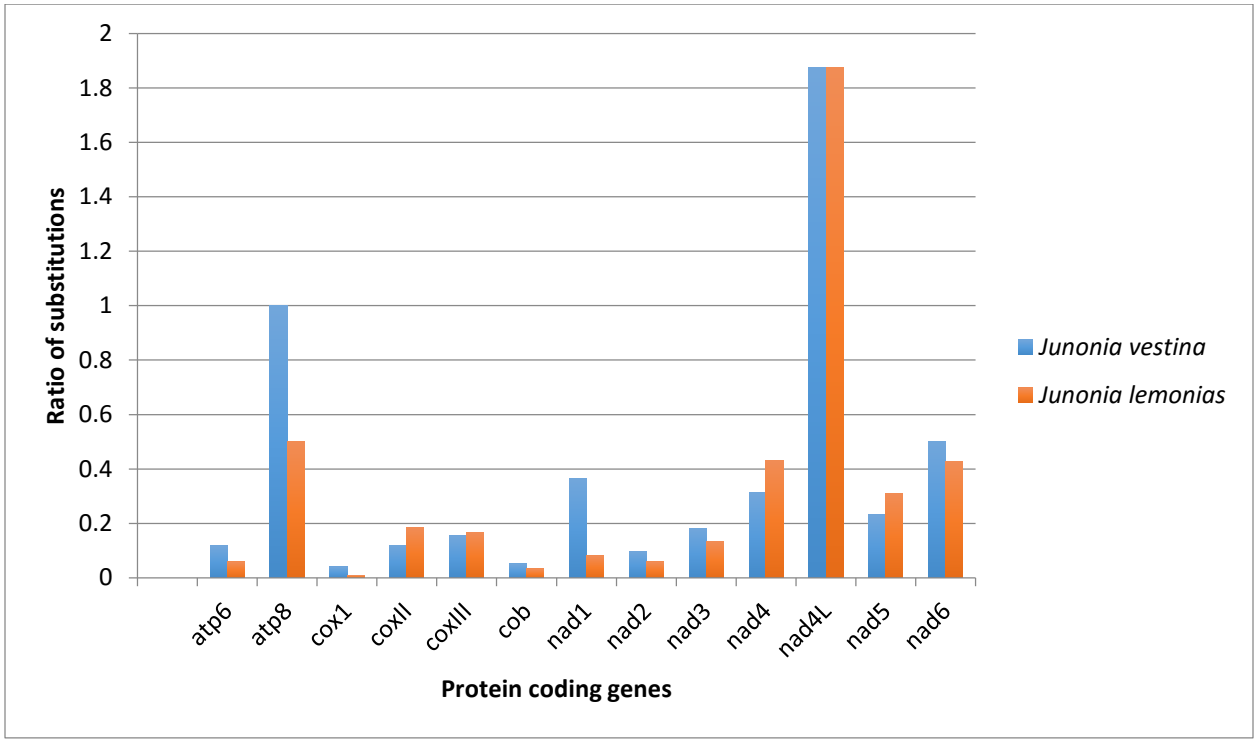
When examining synonymous and non-synonymous mutations in the mitochondrial protein-coding genes, *J. lemonias* had 28 unique non-synonymous sites in mitochondrial protein coding genes, and *J. vestina* had 26 unique non-synonymous changes. There were 46 shared non-synonymous sites between *J. lemonias* and *J. vestina* when compared with *J. orithya madagascariensis* across all mitochondrial protein coding genes. Several genes in *J. vestina* had elevated rates of non-synonymous substitutions relative to the overall substitution rate for the gene, and relative to the rate of non-synonymous substitutions in *J. lemonias* (Fig. 4-1). *Junonia vestina* had one unique non-synonymous site in *atp6*, one in *atp8*, one in *cob*, three in *cox1*, seven in *nad1*, and three in *nad2* **(**Table 4-1). These mutations are potential candidates for high elevation adaptation. The Tajima's D test returned a D value of 5567935.24, indicating there is selection occurring on *J. vestina* and *J. lemonias* relative to *J. orithya madagascariensis* ($p<0.05$ threshold for D for this test = 1.87 (Simonsen et al., 1995)).

For each gene with unique non-synonymous mutations in *J. vestina*, I compared the position of these mutations with high elevation adaptations identified in mammals (Table 4-1).

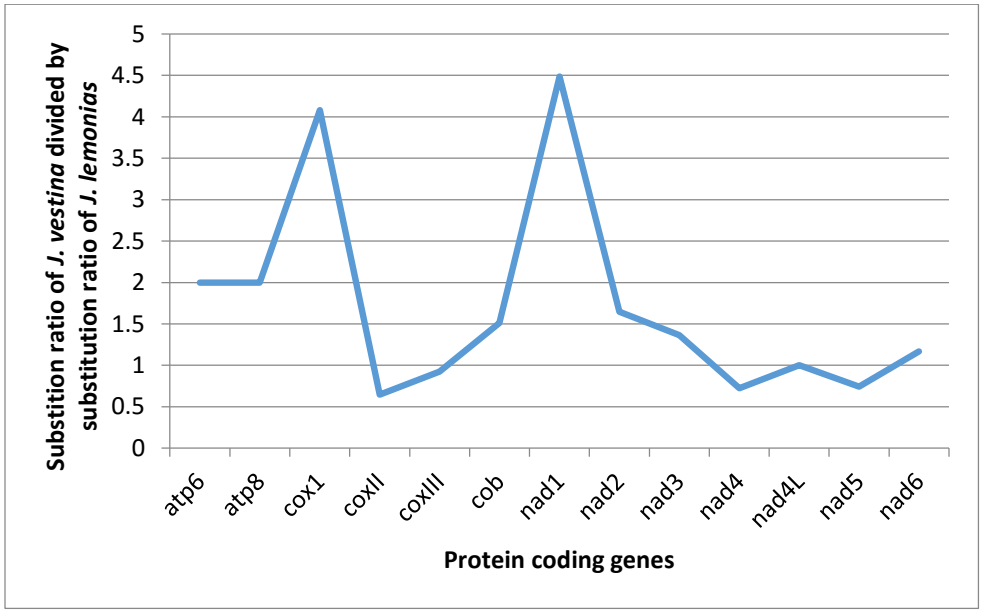
Many of the high elevation species had changes in the same genes. However, no non-synonymous changes at specific sites were shared between them and *J. vestina* (Table 4-1).

Fig. 4-1. A. Ratio of nonsynonymous to synonymous substitutions in the protein coding genes of high elevation *J. vestina* and low elevation sister species *J. lemonias.* B. Ratio of ratios: *J. vestina* ratio of nonsynonymous to synonymous substitutions divided by the *J. lemonias* ratio of nonsynonymous to synonymous substitutions. Genes for which this ratio of ratios was greater than 1.5 show an excess of non-synonymous mutations in *J. vestina* and were considered to be possible targets of selection for adaptation to high elevation: *atp6, atp8, cox1, cob, nad1,* and *nad2*.

A.



B.



There are shared sites of non-synonymous variation between *J. vestina* and *H. sapiens* in *atp6*, *atp8*, *cob*, *cox1*, and *nad1* (Table 4-4). There were associated pathological conditions in *H. sapiens* non-synonymous mutations in all of the identified genes except *cob*. Other sites have

been identified in *H. sapiens* as having non-synonymous variation without any known associated pathology in MITOMAP (Lott et al., 2002).

Discussion

Testing hypotheses for high elevation AT-richness

The AT content of *J. vestina* is higher than other *Junonia* species and may be related to its high elevation environment. The branch length for *J. vestina* appears to be the same as the branch length for *J. lemonias* in the phylogenetic tree based on complete mitochondrial genome sequences (Fig. 3-2). However, it should be noted that *J. lemonias* can complete its life cycle in 25-27 days under optimal conditions (Igarishi & Fukada, 1997) and is active throughout the year (Kunte, 2000), so it likely has 6 or more generations in each calendar year. In contrast, *J. vestina* appears to be bivoltine (2 generations per year) with flight periods in May-July and in September-October (Neild 2008, K. Willmott, pers. comm.). Thus, on a per generation basis, the mutation rate in *J. vestina* is likely much higher than in *J. lemonias*, possibly due to the greater abundance of reactive oxygen species produced by mitochondria at high elevation.

Unlike the high elevation endothermic mammals that have been studied previously (Table 4-1), butterflies are ectotherms. Therefore, elevated AT content in high elevation *Junonia* butterfly species cannot be attributed to higher metabolic rates relative to related species at low elevation. However, butterflies do experience the same combination of low oxygen tension, high UV radiation, and low environmental temperature that have been suggested to produce reactive oxygen species in the mitochondria of high elevation mammals (Hassanin et al., 2009). These results therefore suggest that severe oxidative stress produced by environmental factors (and not elevated metabolic rates) may be sufficient to produce the elevated patterns of AT richness that

has been observed in some high elevation animals. The necessity of maintaining higher metabolic rates for body temperature, and the associated production of reactive oxygen species may also be contributing to this pattern in mammals but it is difficult to separate the two hypotheses from each other due to the linkage of metabolism and high elevation oxidative stress in mammals.

Although AT richness has been observed in some high elevation animals, it is not consistently observed across all high elevation species with sequenced mitochondrial genomes. A study of tribe Caprini (sheep, goats, and relatives) detected elevated AT content in high elevation species, but also in lower-elevation relatives (Hassanin et al., 2009). In contrast, the mitochondrial genome of the high elevation plateau pika (*Ochotona curzoniae*) has reduced AT content relative to related lower-elevation species of the order Lagomorpha with sequenced mitochondrial genomes (Luo et al., 2008). This is also the case in the brine shrimp *Artemia tibetiana* and in the Tibetan stone loach *Triplophysa stoliczkae* —despite occurring at a high elevation, these species have lower AT content than lower-elevation congeners (Wang et al., 2012, Li et al., 2013a, Zhang et al., 2013). Cytosine deaminating to uracil is considered by some authors to be the primary mechanism for producing AT richness in high elevation species (Hassanin et al., 2009) but adenosine can also deaminate under the same conditions to produce a product (hypoxanthine) that is able to base pair similarly to guanine (Zhang et al., 2013). This could explain why some high elevation species actually have elevated GC content, rather than the expected elevated AT content described in Hassanin et al. (2009). Therefore, it may not be elevated AT or GC content per se that represents the mutational signature of high elevation environmental conditions (in both endotherms and ectotherms), but rather an overall elevated mutation rate caused by the reactive oxygen-induced deamination of both cytosine and adenine

in the mitochondrial genomes of animals that live at high altitude. Thus, it appears that there is partial support for hypothesis 1A: high elevation environments produce elevated mutation rates irrespective of metabolic rate.

Mitochondrial protein coding loci under selection at high altitudes

The only other ectothermic species with sequenced mitochondrial genomes in which mitochondrial protein coding loci undergoing adaptive evolution in response to high altitude are the *Triplophysa* loach fishes (Li et al., 2013a), two species of schizothoracine fishes (Li et al., 2013b), and *Artemia* brine shrimp (Zhang et al., 2013). The loaches of the genus *Triplophysa* occur at both high and low elevations in Tibet and other parts of Asia (Kottelat, 2012), which makes *Triplophysa* a useful genus for comparative purposes because it is an ectotherm with low and high elevation species. The loach *Triplophysa stoliczkae* occurs at 5200 meters above sea level in Tibet, making it the highest elevation loach (Kottelat, 2012). However, analyses of the mitochondrial genome of *T. stoliczkae* compared to other members of the genus revealed no non-synonymous mitochondrial changes in its genome. It is possible that insufficient time has elapsed since the divergence for adaptive evolution of protein coding genes to be detectable in *Triplophysa* with sequenced mitochondrial genomes (Li et al., 2013a).

With the mitochondrial genomes of high elevation schizothoracine fish, there was one non-synonymous site identified in *nad5* (Li et al., 2013b) at the same place as a shared non-synonymous site between *J. vestina* and *J. lemonias.* Additional non-synonymous substitutions are found in *atp6, cob, nad1, nad2, nad4,* and other *nad5* locations (Li et al., 2013b), indicating that these genes may play a role in high altitude adaptation. Finally, in the brine shrimp *Artemia*

tibetiana, *atp6, atp8, cox2* and *cox3* have been identified as being related to high altitude adaptation (Zhang et al., 2013).

Although there were no shared non-synonymous sites between *J. vestina* and the high altitude mammalian mitochondrial genomes (Table 4-1), it is still of interest that many of the same genes have been implicated in high altitude adaptation across a variety of species. In particular, *atp6, atp8, cob,* and *nad2* have been implicated repeatedly in high elevation adaptation, including in *J. vestina. cox1* and *nad1* are identified as undergoing adaptive evolution at high elevations less often in the mammal literature, but these two genes have the strongest signals for adaptive evolution in the *Junonia* data set (Fig. 4-1). *atp6, atp8, cox1, cob,* and *nad2* are all part of the mitochondrial oxidative phosphorylation complexes. *atp6* and *atp8* are part of oxidative phosphorylation complex V, *cob* is part of oxidative phosphorylation complex III, *cox1* is part of oxidative phosphorylation complex IV, and *nad 1* and *nad2* are part of oxidative phosphorylation complex I (Moraes et al., 2002). The oxidative phosphorylation pathway is critical in aerobic energy production and helps maintain body temperature in endotherms. Mutations in these genes may be involved in fine-tuning efficient energy production under high elevation conditions, and the candidate positions for adaptive evolution seem to be distributed across nearly all of the complexes that make up the mitochondrial electron transport chain. These mutations may be adaptive at high elevation, even if they are deleterious under other environmental conditions. Thus, there appears to be support for hypothesis 2A (adaptive evolution to high elevation conditions in the mitochondrial genome of *J. vestina* butterflies) and hypothesis 3A (parallel evolution among the mitochondrial genomes of high elevation animals).

All of the non-synonymous sites in *J. vestina* along with their parallel sites in *H. sapiens* and the location of the mutations in the protein domains are listed in Table 4-5. Nine of the

unique non-synonymous sites in mitochondrial protein coding genes in *J. vestina* are in transmembrane domains, four are in domains in the mitochondrial matrix, and three are located in domains in the inter-membrane space. Many of the transmembrane sites line the ion channels created by transmembrane mitochondrial proteins and through which H+ ions flow through. Changes in the amino acids lining these channels could affect the rate at which H+ ions flow through the channel. Among the other non-synonymous mutations discovered in *J. vestina,* the most potentially significant in terms of mitochondrial function is in *cox1* (position G136S in *J. vestina*, P139S in *H. sapiens),* close to the active site of the protein (Rimon et al., 2010) . The presence of a serine residue rather than a glycine in *J. vestina* is noteworthy because serine has a greater ability than glycine to form hydrogen bonds with oxygen molecules (Berg et al. 2007). The mutation to serine in *J. vestina* could facilitate pre-loading oxygen molecules into the active site of *cox1* and may facilitate the use of oxygen as the final electron receptor for aerobic respiration under conditions of low oxygen tension.

Table 4-5. Amino acid positions of non-synonymous sites along with where these amino acids occur within the protein.

Gene	***J. vestina* AA position (*J. orithya* AA, position number, *J. vestina* AA)**	***H. sapiens* AA position (wildtype AA, position number)**	**Protein domain**
atp6	S64T	T63	1st mitochondrial matrix domain
atp8	M35I	M42	1st mitochondrial matrix domain
cob	L332M	L328	8th transmembrane domain
cox1	G136S	P139	2nd mitochondrial matrix domain
	I466V	I469	12th transmembrane domain
nad1	I9V	P2	1st transmembrane domain
	I11V	A4	1st transmembrane domain
	V77I	L70	2nd transmembrane domain
	M104T	N97	2nd intermembrane domain
	I109V	L102	4th transmembrane domain
	V177I	E170	3rd intermembrane domain
	G312D	S312	8th transmembrane domain
nad2	A26T	S23	1st transmembrane domain
	N98D	M99	3rd transmembrane domain
	V220I	S223	4th mitochondrial matrix domain

Parallels between human variation and mitochondrial adaptation in high elevation species

While there were no specific parallels between high elevation adaptation in human populations and high elevation adaptation in *J. vestina* butterflies (Table 4-1), I have been able to identify naturally occurring human variation at lower elevations at homologous positions to possible adaptive changes in *J. vestina* (Table 4-4). In some cases, the associated pathological conditions in the human mitochondrial protein coding genes provide insights into the possible functional consequences of substitutions in high elevation species. Many mitochondrial variants have been identified in humans in the process of investigating disease. The M42T (T8490C) variant was investigated for a potential role in multiple sclerosis, although it was ultimately not implicated as a causative agent for this illness (Ban et al., 2008). Similarly, the amino acid substitution P139S in *cox1* in humans (homologous to amino acid substitution G136S in *J. vestina* discussed in the previous section) was investigated as a potential factor in the flow of calcium ions in human patients (Kazuno et al., 2006). Some of the amino acid changes that occur in *J. vestina* and humans are associated in humans with multiple diseases, such as the A4T (G3316A) change in *nad1* that has been associated with Leber hereditary optic neuropathy and megakaryoblastic leukaemic cells (Mostafaie et al., 2004, Piccoli et al., 2008). Some of these variants have been discovered within groups that do have pathological conditions, yet may not cause structural changes to proteins (Lehtonen et al., 2003) or may require other factors to cause disease symptoms (Puomila et al., 2007). Due to the phenotypes associated with each of these mutations in humans, these amino acid positions appear to have functional consequences for the activity of the proteins in humans. The parallel mutations observed at homologous positions within the *J. vestina* mitochondrial genome represent the best candidates for functional sequence positions that have undergone adaptive evolution in response to high altitude conditions.

There are additional non-synonymous variants that have also been identified in human mitochondrial genomes with no identified associated pathologies in MITOMAP (Lott et al., 2002). These variants may not have discernable human phenotypes because they have only been evaluated at low elevation, or they may be entirely selectively neutral variants. They represent an area of further study to determine if there are pathological or other phenotypic conditions yet to be identified at these sequence variant locations, but based on currently available information, they are less likely to have functional consequences in high elevation *J. vestina* butterflies.

Conclusions

Evaluation of the complete mitochondrial genome of *J. vestina* butterflies in the context of high elevation adaptation provides insights into the mechanisms underlying the elevated mutation rates that exist in high altitude species. Because *J. vestina* is an ectotherm, yet shows the same elevated mutation rates of well-studied high elevation endotherms, one can conclude that environmental factors such as low oxygen concentration, high UV radiation, and low temperatures are sufficient to produce high rates of reactive oxygen species. Changes in metabolic rate do not need to be invoked to explain increases in the frequency of nucleotide deamination and consequent increases in the rate of sequence evolution in the mitochondrial genomes of high elevation animals.

This analysis has also allowed us to identify a set of protein-coding mitochondrial loci (*atp6, atp8, cox1, cob,* and *nad2*) that repeatedly seem to be under selection in response to high elevation conditions. While the genes implicated are consistent, the candidate positions for adaptive evolution within the genes are usually not the same in different high elevation species. However, many of the candidate positions within mitochondrial loci identified in *J. vestina* do

echo variation that is observed in the human population, some of which is associated with pathological conditions and which may help associate function with high altitude sequence variation. As the cost of mitochondrial genome sequencing continues to decrease due to new sequencing technologies, the number of high altitude species with sequenced mitochondrial genomes will also continue to grow. Using diverse organisms such as mammals and butterflies to study this phenomenon provides new opportunities for testing the generality of the trends identified for mitochondrial genome evolution at high elevation.

Literature Cited

Balloux, F., Handley, L.-J.L., Jombart, T., Liu, H. and Manica, A., 2009. Climate shaped the worldwide distribution of human mitochondrial DNA sequence variation. Proc. R. Soc. Lond., B: Biol. Sci.

Ban, M., Elson, J., Walton, A., Turnbull, D., Compston, A., Chinnery, P. and Sawcer, S., 2008. Investigation of the role of mitochondrial DNA in Multiple Sclerosis susceptibility. PLoS ONE 3, e2891.

Berg, J.M., Tymoczko, J.L. and Stryer, L., 2007. Biochemistry, W. H. Freeman and Company, New York.

Chen, J., Hattori, Y., Nakajima, K., Eizawa, T., Ehara, T., Koyama, M., Hirai, T., Fukuda, Y., Kinoshita, M., Sugiyama, A., Hayashi, J.-I., Onaya, T., Kobayashi, T. and Tawata, M., 2006. Mitochondrial complex I activity is significantly decreased in a patient with maternally inherited type 2 diabetes mellitus and hypertrophic cardiomyopathy associated with mitochondrial DNA C3310T mutation: A cybrid study. Diabetes Res. Clin. Pract. 74, 148-153.

EndMemo, 2015. DNA/RNA GC Content Calculator. EndMemo.

Finch, T.M., Zhao, N., Korkin, D., Frederick, K.H. and Eggert, L.S., 2014. Evidence of Positive Selection in Mitochondrial Complexes I and V of the African Elephant. PLoS ONE 9, e92587.

Foote, A.D., Morin, P.A., Durban, J.W., Pitman, R.L., Wade, P., Willerslev, E., Gilbert, M.T.P. and da Fonseca, R.R., 2010. Positive selection on the killer whale mitogenome. Biology Letters.

Gemmell, A.P. and Marcus, J.M., 2015. A tale of two haplotype groups: The origin and distribution of divergent New World *Junonia COI* haplotypes. Syst. Ent. 40, 532-546. doi: 10.1111/syen.12120.

Gu, M., Dong, X., Shi, L., Shi, L., Lin, K., Huang, X. and Chu, J., 2012. Differences in mtDNA whole sequence between Tibetan and Han populations suggesting adaptive selection to high altitude. Gene 496, 37-44.

Hassanin, A., Ropiquet, A., Coulous, A. and Cruaud, C., 2009. Evolution of the mitochondrial genome in mammals living at high altitude: New insights from a study of the tribe

Caprini (Bovidae, Antilopinae). J. Mol. Evol. 68, 293-310 DOI 10.1007/s00239-009-9208-7.
Hattori, Y., Nakajima, K., Eizawa, T., Ehara, T., Koyama, M., Hirai, T., Fukuda, Y. and Kinoshita, M., 2003. Heteroplasmic mitochondrial DNA 3310 mutation in NADH Dehydrogenase Subunit 1 associated with Type 2 Diabetes, Hypertrophic Cardiomyopathy, and Mental Retardation in a single patient. Diabetes Care 26, 952-953.
Hattori, Y., Takeoka, M., Nakajima, K., Ehara, T. and Koyama, M., 2005. A heteroplasmic mitochondrial DNA 3310 mutation in the ND1 gene in a patient with Type 2 Diabetes, Hypertrophic Cardiomyopathy, and Mental Retardation. Exp. Clin. Endocrinol Diabetes 113, 318-323.
Igarishi, S. and Fukada, H., 1997. The Life Histories of Asian Butterflies, Volume 1, Tokai University Press.
Irwin, D.E., 2012. Local adaptation along smooth ecological gradients causes phylogeographic breaks and phenotypic clustering. Am. Nat. 180, 35-49 DOI: 10.1086/666002.
Kang, L., Zheng, H.-X., Chen, F., Yan, S., Liu, K., Qin, Z., Liu, L., Zhao, Z., Li, L., Wang, X., He, Y. and Jin, L., 2013. mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands. Mol. Biol. Evol. 30, 2579-2587.
Kazuno, A.-a., Munakata, K., Nagai, T., Shimozono, S., Tanaka, M., Yoneda, M., Kato, N., Miyawaki, A. and Kato, T., 2006. Identification of mitochondrial DNA polymorphisms that alter mitochondrial matrix pH and intracellular calcium dynamics. PLoS Genetics 2, e128.
Kleckova, I., Konvicka, M. and Klecka, J., 2014. Thermoregulation and microhabitat use in mountain butterflies of the genus *Erebia*: Importance of fine-scale habitat heterogeneity. Journal of Thermal Biology 41, 50-58.
Kodandaramaiah, U., 2009. Eyespot evolution: Phylogenetic insights from *Junonia* and related butterfly genera (Nymphalidae: Junoniini). Evol. Dev. 11, 489-497.
Kodandaramaiah, U. and Wahlberg, N., 2007. Out-of-Africa origin and dispersal-mediated diversification of the butterfly genus *Junonia* (Nymphalidae: Nymphalinae). J. Evol. Biol. 20, 2181-2191.
Kottelat, M., 2012. Conspectus cobitidum*: an inventory of the loaches of the world (Teleostei: Cypriniformes: Cobitoidei). Raffles Bull. Zool. Supplement No. 26, 1-199.
Kow, Y.W., 2002. Repair of deaminated bases in DNA. Free Radic. Biol. Med. 33, 886-893.
Kunte, K., 2000. Butterflies of Peninsular India, Universities Press, Hyderabad, India.
Lehtonen, M., Moilanen, J. and Majamaa, K., 2003. Increased variation in mtDNA in patients with familial sensorineural hearing impairment. Hum. Genet. 113, 220-227.
Li, J., Si, S., Guo, R., Wang, Y. and Song, Z., 2013a. Complete mitochondrial genome of the stone loach, *Triplophysa stoliczkae* (Teleostei: Cypriniformes: Balitoridae). Mitochondrial DNA 24, 8-10.
Li, Y., Ren, Z., Shedlock, A.M., Wu, J., Sang, L., Tersing, T., Hasegawa, M., Yonezawa, T. and Zhong, Y., 2013b. High altitude adaptation of the schizothoracine fishes (Cyprinidae) revealed by the mitochondrial genome analyses. Gene 517, 169-178.
Liu, C., Yang, Q., Hwang, S.-J., Sun, F., Johnson, A.D., Shirihai, O.S., Vasan, R.S., Levy, D. and Schwartz, F., 2012. Association of genetic variation in the mitochondrial genome with blood pressure and metabolic traits. Hypertension 60, 949-956.
Lobry, J.R., 1996. Asymmetric substitution patterns in the two DNA strands of bacteria. Mol. Biol. Evol. 13, 660-665.

Lott, M.T., Leipzig, J.N., Derbeneva, O., Xie, H.M., Chalkia, D., Sarmady, M., Procaccio, V. and Wallace, D.C., 2002. mtDNA Variation and Analysis Using Mitomap and Mitomaster. Curr. Protoc. Bioinformatics. John Wiley & Sons, Inc.

Lu, J., Qian, Y., Li, Z., Yang, A., Zhu, Y., Li, R., Yang, L., Tang, X., Chen, B., Ding, Y., Li, Y., You, J., Zheng, J., Tao, Z., Zhao, F., Wang, J., Sun, D., Zhao, J., Meng, Y. and Guan, M.-X., 2010. Mitochondrial haplotypes may modulate the phenotypic manifestation of the deafness-associated 12S rRNA 1555A>G mutation. Mitochondrion 10, 69-81.

Luo, Y., Chen, Y., Liu, F. and Gao, Y., 2012. Mitochondrial genome of Tibetan wild ass (*Equus kiang*) reveals substitutions in NADH which may reflect evolutionary adaptation to cold and hypoxic conditions. Asia Life Sci. 21, 1-11.

Luo, Y., Gao, W., Gao, Y., Tang, S., Hugan, Q., Tan, X., Chen, J. and Huang, T., 2008. Mitochondrial genome analysis of *Ochotona curzoniae* and implication of cytochrome c oxidase in hypoxic adaptation. Mitochondrion 8, 352-357 doi:10.1016/j.mito.2008.07.005.

Luo, Y., Yang, X. and Gao, Y., 2013. Mitochondrial DNA response to high altitude: a new perspective on high-altitude adaptation. Mitochondrial DNA 24, 313-319. doi: 10.3109/19401736.2012.760558.

Martin, A.P. and Palumbi, S.R., 1993. Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. U.S.A. 90, 4087-4091.

McCullagh, B.S. and Marcus, J.M., 2015. The complete mitochondrial genome of the lemon pansy *Junonia lemonias* (Lepidoptera: Nymphalidae: Nymphalinae). J. Asia Pac. Entomol. 18, 749-755.

Moraes, C.T., Srivastava, S., Kirkinezos, I., Oca-Cossio, J., vanWaveren, C., Woischnick, M. and Diaz, F., 2002. Mitochondrial DNA structure and function. Int. Rev. Neurobiol. pp. 3-23. Academic Press.

Mostafaie, N., Rossmanith, W., Hombauer, H., Dechat, T., Raffelsberger, T., Bauer, K., Worofka, B., Kittl, E., Hofmann, J., Hejtman, M., Kirchmeyr, W., Schreiber, W., Weissgram, S., Jungwirth, S., Fischer, P., Bittner, R. and Huber, K., 2004. Mitochondrial genotype and risk for Alzheimer's disease: cross-sectional data from the Vienna-Transdanube-Aging "VITA" study. J. Neural Transm. 111, 1155-1165.

Neild, A.F.E., 2008. The Butterflies of Venezuela, Part 2: Nymphalidae II (Acraeinae, Libytheinae, Nymphalinae, Ithomiinae, Morphinae), Meridian, London.

Pfeiler, E., Johnson, S. and Markow, T.A., 2012. DNA barcodes and insights into the relationships and systematics of buckeye butterflies (Nymphalidae: Nymphalinae: *Junonia*) from the Americas. J. Lepid. Soc. 66, 185-198.

Piccoli, C., Ripoli, M., Scrima, R., Stanziale, P., Di Ianni, M., Moretti, L., Biscottini, B., Carella, M., Boffoli, D., Tabilio, A. and Capitanio, N., 2008. MtDNA mutation associated with mitochondrial dysfunction in megakaryoblastic leukaemic cells. Leukemia 22, 1938-1941.

Puomila, A., Hamalainen, P., Kivioja, S., Savontaus, M.-L., Koivumaki, S., Huoponen, K. and Nikoskelainen, E., 2007. Epidemiology and penetrance of Leber hereditary optic neuropathy in Finland. Eur. J. Hum. Genet. 15, 1079-1089.

Qu, J., Zhou, X., Zhao, F., Liu, X., Zhang, M., Sun, Y.-H., Liang, M., Yuan, M., Liu, Q., Tong, Y., Wei, Q.-P., Yang, L. and Guan, M.-X., 2010. Low penetrance of Leber's hereditary optic neuropathy in ten Han Chinese families carrying the ND6 T11484C mutation. Biochim. Biophys. Acta 1800, 305-312.

Rimon, G., Sidhu, R.S., Lauver, D.A., Lee, J.Y., Sharma, N.P., Yuan, C., Frieler, R.A., Trievel, R.C., Lucchesi, B.R. and Smith, W.L., 2010. Coxibs interfere with the action of aspirin by binding tightly to one monomer of cyclooxygenase-1. Proc. Natl. Acad. Sci. 107, 28-33.

Shi, Q., Huang, D., Wang, Y. and Hao, J., 2013. The complete mitochondrial genome of Blue Pansy, *Junonia orithya* (Lepidoptera: Nymphalidae: Nymphalinae). Mitochondrial DNA 0, 1-2. doi: 10.3109/19401736.19402013.19823182.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D. and Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7.

Simonsen, K.L., Churchill, G.A. and Aquadro, C.F., 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141, 413-29.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585-95.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30, 2725-2729.

Tang, D.-L., Zhou, X., Li, X., Zhao, L. and Liu, F., 2006. Variation of mitochondrial gene and the association with type 2 diabetes mellitus in a Chinese population. Diabetes Res. Clin. Pract. 73, 77-82.

Tang, S., Wang, J., Zhang, V.W., Li, F.-Y., Landsverk, M., Cui, H., Truong, C.K., Wang, G., Chen, L.C., Graham, B., Scaglia, F., Schmitt, E.S., Craigen, W.J. and Wong, L.-J.C., 2013. Transition to next generation analysis of the whole mitochondrial genome: A summary of molecular defects. Hum. Mutat. 34, 882-893.

Torroni, A., Miller, J.A., Moore, L.G., Zamudio, S., Zhuang, J., Droma, T. and Wallace, D.C., 2005. Mitochondrial DNA analysis in Tibet: Implications for the origin of the Tibetan population and its adaptation to high altitude. Am. J. Phys. Anthropology 93, 189-199.

Wang, J., Tang, Q., Wang, Z., Zhang, Y., Wu, Q. and Peng, Z., 2012. The complete mitogenome sequence of a cave loach *Triplophysa rosa* (Teleostei, Balitoridae, Nemacheilinae). Mitochondrial DNA 23, 366-368.

Wu, L.-W., Lin, L.-H., Lees, D. and Hsu, Y.-F., 2014. Mitogenomic sequences effectively recover relationships within brush-footed butterflies (Lepidoptera: Nymphalidae). BMC Genomics 15, 468 doi:10.1186/1471-2164-15-468.

Xu, S., Luosang, J., Hua, S., He, J., Ciren, A., Wang, W., Tong, X., Liang, Y., Wang, J. and Zheng, X., 2007. High altitude adaptation and phylogenetic analysis of Tibetan horse based on the mitochondrial genome. J. Genet. Genomics 34, 720-729.

Xu, S.Q., Yang, Y.Z., Zhou, J., Jing, G.E., Chen, Y.T., Wang, J., Yang, H.M., Wang, J., Yu, J., Zheng, X.G. and Ge, R.L., 2005. A mitochondrial genome sequence of the Tibetan antelope (*Pantholops hodgsonii*). Genomics, proteomics & bioinformatics 3, 5-17.

Yao, Y.-G., Kong, Q.-P., Salas, A. and Bandelt, H.-J., 2008. Pseudomitochondrial genome haunts disease studies. J. Med. Genet. 45, 769-772.

Yu, L., Wang, X., Ting, N. and Zhang, Y., 2011. Mitogenomic analysis of Chinese snub-nosed monkeys: Evidence of positive selection in NADH dehydrogenase genes in high-altitude adaptation. Mitochondrion 11, 497-503.

Zhang, H., Luo, Q., Sun, J., Liu, F., Wu, G., Yu, J. and Wang, W., 2013. Mitochondrial genome sequences of *Artemia tibetiana* and *Artemia urmiana*: assessing molecular changes for high plateau adaptation. Sci. China Life Sci. 56, 440-452.
Zhao, X., Wu, N., Zhu, Q., Gaur, U., Gu, T. and Li, D., 2015. High-altitude adaptation of Tibetan chicken from MT-COI and ATP-6 perspective. Mitochondrial DNA, 1-9.

Chapter 5: Conclusion

Next generation sequencing of mitochondrial genomes has been helpful for clarifying aspects of the phylogeny of *Junonia* butterflies. In Chapter 2, *Junonia lemonias*, the lemon pansy butterfly, was shown to have a mitochondrial genome of 15,230 bp. The mitochondrial genome contains 13 protein coding genes, 22 tRNAs, and 2 rRNAs. The control region is 333 bp long. The *J. lemonias* mitogenome has 80.40% AT content. The general arrangement of the *J. lemonias* mitogenome is the same as the other two sequenced *Junonia* mitochondrial genomes (Shi et al., 2013, Wu et al., 2014). More broadly, it maintains synteny with other species of Nymphalid butterflies. The synteny of *J. lemonias* with other Nymphalid butterflies in general and with other *Junonia* species in particular allows whole mitochondrial genomes to be aligned to produce sets of homologous characters for phylogenetic analysis. This was consistent with prior work exploring Nymphalid butterfly genomes (Timmermans et al., 2014, Wu et al., 2014).

Phylogenetic analysis of Nympahlid mitochondrial genomes including *J. lemonias* shows that the subfamily Nymphalinae, of which *Junonia* is a part, is monophyletic. In the analysis in Chapter 2, which included three *Junonia* mitochondrial genomes, *Junonia* was recovered as a monophyletic group. When using only protein coding sequences from the mitochondrial genomes, *Junonia* was still monophyletic. Considering that *J. lemonias* has synteny with other Nymphalids and that *Junonia* is monophyletic using whole mitochondrial genomes, this means that the sequencing and analysis of whole mitochondrial genomes are an effective strategy for investigating *Junonia* phylogeny. The closest outgroups for genus *Junonia* with available mitochondrial genome sequences are *Yoma sabina* and *Hypolimnas bolina*, which is consistent with prior analyses of smaller molecular data sets (Kodandaramaiah & Wahlberg, 2007, Kodandaramaiah, 2009) and which were used for further phylogenetic tree construction focusing on the New World *Junonia*.

Phylogenetic analysis of the 16 mitochondrial genomes that I sequenced myself and the 2 *Junonia* mitochondrial genomes and appropriate outgroups sequenced by others (Shi et al., 2013, Wu et al., 2014, Shi et al., 2015) produced a robust phylogeny of *Junonia* (Chapter 3) that differs from previous phylogenetic hypotheses based of *cox1* barcodes (Brévignon & Brévignon, 2012, Pfeiler et al., 2012, Gemmell et al., 2014, Gemmell & Marcus, 2015). The clades in the mitogenome tree were well supported with both maximum likelihood and parsimony bootstrap. The use of full length mitochondrial genome sequences was able to strongly resolve relationships unable to be resolved by *cox1* DNA barcodes (Gemmell & Marcus, 2015).

Almost all prior taxonomic authorities have hypothesized that the New World *Junonia* have been monophyletic and descended from a single colonization event from the Old World (Forbes, 1928, Eliot, 1946, Eliot, 1947a, Eliot, 1947b, Forbes, 1947, Corbet, 1948, Eliot, 1949, Kodandaramaiah & Wahlberg, 2007, Kodandaramaiah, 2009, Brévignon & Brévignon, 2012, Pfeiler et al., 2012, Borchers & Marcus, 2014, Gemmell et al., 2014). Authorities have differed on whether *Junonia* arrived from Africa or the Indo-Pacific, and which Old World species is the sister clade to the New World diversification (Kodandaramaiah & Wahlberg, 2007). Only one prior study has suggested that *Junonia* arrived in the New World had multiple invasion events (Gemmell & Marcus, 2015). From analysis of complete mitochondrial genome sequences (Chapter 3), it appears that there were three separate invasion events of *Junonia* species to the New World. Haplotypes A2 and B diverged from each other approximately 1.79+/-0.34 and 2.31+/-0.42 million years ago, and New World haplotype group A2 is more closely related to extant Australian *J. villida* mitochondrial haplotypes that it is to New World haplotype group B. However, both haplotype groups A2 and B are both closely related to the *J. villida* mitochondrial lineage. Haplotype A2 was in the New World and began to diversify by approximately 0.96+/-

0.29 million years ago, whereas haplotype B was in the New World and diversifying by approximately 1.18+/-0.29 to 1.19+/-0.29 million years ago. Further, haplotype groups A2 and B are both found in most New World *Junonia* species, suggesting that the descendants of the waves of immigration that established these haplotype groups in the New World probably hybridized with one another once they arrived.

Haplotype A1 that is found only in *J. vestina* diverged from *J. lemonias*, its Asian sister clade, about 1.60+/-0.32 million years ago, meaning that *J. vestina* could not have been present in the New World before that date. *Junonia vestina* not only has a distinct A1 mitochondrial haplotype (Pfeiler et al., 2012), but I determined that it also has several mitochondrial protein coding genes that appear to be high altitude adapted (Chapter 4). Potential high altitude adaptations were found in multiple genes: *atp6, atp8, cox1, cob, nad1,* and *nad2*. The genes in *J. vestina* with the most elevated nonsynonymous mutation rates were *cox1* and *nad1*. Many of these genes were implicated as potentially being linked to high altitude adaptation in other high elevation species (Luo et al., 2013). When checking the homologous positions in human mitochondrial genomes using MitoMap (Lott et al., 2002), several shared sites of non-synonymous variation were identified. Some of these sites in humans have links to diseases, such as type 2 diabetes (Hattori et al., 2003, Hattori et al., 2005), Leber hereditary optic neuropathy (Puomila et al., 2007, Qu et al., 2010), and multiple sclerosis (Ban et al., 2008). *Cox1* and *nad1* are both part of the oxidative phosphorylation complex (Moraes et al., 2002), and therefore the nonsynonymous mutations could be related to adaptation for low atmospheric oxygen concentration. A further benefit of the examination of the *J. vestina* mitochondrial genome is that unlike almost all other high elevation species studied for high elevation adaption mitochondrial

adaptation, *J. vestina* is an ectotherm, so the changes and adaptations occurring are not related to endothermy.

Future Directions

As next generation sequencing costs continue to decrease, it is to be expected that more laboratories will employ this method. To produce a more robust *Junonia* phylogeny, there should be an effort to sequence the mitochondrial genomes of all known *Junonia* species. Given that there are approximately 29 *Junonia* species in the Old and New Worlds, and that there are now full-length mitochondrial genomes for 16 different *Junonia* samples, it should be relatively easy to sequence the remaining species. The most prominent difficulties may arise from acquiring well preserved specimens of the rarer *Junonia* species to sequence, but this could potentially be solved by collaboration between labs with different access to materials. Acquiring mitochondrial genome sequences from additional specimens of widespread species such as *J. orithya* and *J. villida* may help clarify the relationships among populations that are not monophyletic in some phylogenetic analyses (Fig. 3-1) and also narrow the windows during which *Junonia* colonization of the New World took place.

Although mitochondrial genome sequences are very useful, it would also be of interest to have more sequences from the nuclear genome of *Junonia* species. Given that mitochondrial genomes and nuclear genomes evolve at different rates, and are inherited by different mechanisms (maternal vs. biparental) comparing nuclear and mitochondrial genome-based phylogenies would be very interesting. Partial sequences from two nuclear genes have already been used for phylogenetic analysis in *Junonia*, (Kodandaramaiah & Wahlberg, 2007), but they are of modest length and were not analyzed separately from mitochondrial sequences. It would be useful to know what full length nuclear DNA sequences can tell us about phylogeny, as they

may better reflect the relationships among *Junonia* species, independent of introgression of mitochondrial haplotypes. Finally, it would be potentially very informative to acquire a complete genome sequence from a *Junonia* species. This would open up many new possibilities for the study of diversification, adaptation, and hybridization in this genus.

Literature Cited

Ban, M., Elson, J., Walton, A., Turnbull, D., Compston, A., Chinnery, P. and Sawcer, S., 2008. Investigation of the role of mitochondrial DNA in Multiple Sclerosis susceptibility. PLoS ONE 3, e2891.

Borchers, T.E. and Marcus, J.M., 2014. Genetic population structure of buckeye butterflies (*Junonia*) from Argentina. Syst. Ent. 39, 242-255. doi: 10.1111/syen.12053.

Brévignon, L. and Brévignon, C., 2012. Nouvelles observations sur le genre *Junonia* en Guyane Française. (Lepidoptera: Nymphalidae) (Seconde partie). Lépidoptères de Guyane 7, 8-35.

Corbet, A.S., 1948. Papers on Malaysian Rhopalocera. V. The conspecificity of the American *Precis lavinia* (Cramer) with the Oriental *Precis orithya* (Cramer). Entomologist 81, 54-56.

Eliot, N., 1946. Continental drift and *Precis lavinia.* Entomologist 79, 225-228.

Eliot, N., 1947a. More on continental drift, *Precis lavinia* Hb. and *P. villida* F. Entomologist 80, 230-234.

Eliot, N., 1947b. More on continental drift, *Precis lavinia* Hb. and *P. villida* F.: A postscript. Entomologist 80, 283.

Eliot, N., 1949. The significance of wing pattern in *Precis* (Nymphalidae) in America. Entomologist 82, 176-184.

Forbes, W.T.M., 1928. Variation in *Junonia lavinia* (Lepidoptera, Nymphalidae). J. N.Y. Ent. Soc. 36, 306-321.

Forbes, W.T.M., 1947. Buckeyes and Wegener. Entomologist 80, 56-58.

Gemmell, A.P., Borchers, T.E. and Marcus, J.M., 2014. Genetic Population Structure of Buckeye Butterflies (*Junonia*) from French Guiana, Martinique, and Guadeloupe. Psyche 2014, 1-21. doi: 10.1155/2014/897596.

Gemmell, A.P. and Marcus, J.M., 2015. A tale of two haplotype groups: The origin and distribution of divergent New World *Junonia COI* haplotypes. Syst. Ent. 40, 532-546. doi: 10.1111/syen.12120.

Hattori, Y., Nakajima, K., Eizawa, T., Ehara, T., Koyama, M., Hirai, T., Fukuda, Y. and Kinoshita, M., 2003. Heteroplasmic mitochondrial DNA 3310 mutation in NADH Dehydrogenase Subunit 1 associated with Type 2 Diabetes, Hypertrophic Cardiomyopathy, and Mental Retardation in a single patient. Diabetes Care 26, 952-953.

Hattori, Y., Takeoka, M., Nakajima, K., Ehara, T. and Koyama, M., 2005. A heteroplasmic mitochondrial DNA 3310 mutation in the ND1 gene in a patient with Type 2 Diabetes, Hypertrophic Cardiomyopathy, and Mental Retardation. Exp. Clin. Endocrinol Diabetes 113, 318-323.

Kodandaramaiah, U., 2009. Eyespot evolution: Phylogenetic insights from *Junonia* and related butterfly genera (Nymphalidae: Junoniini). Evol. Dev. 11, 489-497.

Kodandaramaiah, U. and Wahlberg, N., 2007. Out-of-Africa origin and dispersal-mediated diversification of the butterfly genus *Junonia* (Nymphalidae: Nymphalinae). J. Evol. Biol. 20, 2181-2191.

Lott, M.T., Leipzig, J.N., Derbeneva, O., Xie, H.M., Chalkia, D., Sarmady, M., Procaccio, V. and Wallace, D.C., 2002. mtDNA Variation and Analysis Using Mitomap and Mitomaster. Curr. Protoc. Bioinformatics. John Wiley & Sons, Inc.

Luo, Y., Yang, X. and Gao, Y., 2013. Mitochondrial DNA response to high altitude: a new perspective on high-altitude adaptation. Mitochondrial DNA 24, 313-319. doi: 10.3109/19401736.2012.760558.

Moraes, C.T., Srivastava, S., Kirkinezos, I., Oca-Cossio, J., vanWaveren, C., Woischnick, M. and Diaz, F., 2002. Mitochondrial DNA structure and function. Int. Rev. Neurobiol. pp. 3-23. Academic Press.

Pfeiler, E., Johnson, S. and Markow, T.A., 2012. DNA barcodes and insights into the relationships and systematics of buckeye butterflies (Nymphalidae: Nymphalinae: *Junonia*) from the Americas. J. Lepid. Soc. 66, 185-198.

Puomila, A., Hamalainen, P., Kivioja, S., Savontaus, M.-L., Koivumaki, S., Huoponen, K. and Nikoskelainen, E., 2007. Epidemiology and penetrance of Leber hereditary optic neuropathy in Finland. Eur. J. Hum. Genet. 15, 1079-1089.

Qu, J., Zhou, X., Zhao, F., Liu, X., Zhang, M., Sun, Y.-H., Liang, M., Yuan, M., Liu, Q., Tong, Y., Wei, Q.-P., Yang, L. and Guan, M.-X., 2010. Low penetrance of Leber's hereditary optic neuropathy in ten Han Chinese families carrying the ND6 T11484C mutation. Biochim. Biophys. Acta 1800, 305-312.

Shi, Q.-H., Sun, X.-Y., Wang, Y.-L., Hao, J.-S. and Yang, Q., 2015. Morphological Characters Are Compatible with Mitogenomic Data in Resolving the Phylogeny of Nymphalid Butterflies (Lepidoptera: Papilionoidea: Nymphalidae). PLoS ONE 10.

Shi, Q., Huang, D., Wang, Y. and Hao, J., 2013. The complete mitochondrial genome of Blue Pansy, *Junonia orithya* (Lepidoptera: Nymphalidae: Nymphalinae). Mitochondrial DNA 0, 1-2. doi: 10.3109/19401736.19402013.19823182.

Timmermans, M.J.T.N., Lees, D.C. and Simonsen, T.J., 2014. Towards a mitogenomic phylogeny of Lepidoptera. Mol. Phylogenet. Evol. 79, 169-178.

Wu, L.-W., Lin, L.-H., Lees, D. and Hsu, Y.-F., 2014. Mitogenomic sequences effectively recover relationships within brush-footed butterflies (Lepidoptera: Nymphalidae). BMC Genomics 15, 468 doi:10.1186/1471-2164-15-468.